by Anhalt University of Applied Sciences

Proceedings of the 11th International Conference on Applied Innovations in IT

Volume 11

Issue 2

EDITION
Hochschule Anhalt

Proceedings of the 11th International Conference on Applied Innovations in IT

Volume 11 | Issue 2

Koethen, Germany
30 November 2023

This journal is devoted to problems of applied research in the fields of automation, communication and information technologies. The research results are of interest for researchers and development engineers who deal with theoretical base and the application of the knowledge in the respective areas.

ISBN: 978-3-96057-171-1 (Online)
ISSN: 2199-8876

All papers published in this jornal are full-open access:
immediately freely available to read and download.

Publisher: Anhalt University of Applied Sciences
Bernburg / Koethen / Dessau
http://www.hs-anhalt.de

Postfach 1458
D-06354 Koethen, Germany
Phone: +49 3496 67 2327
Email: eduard.siemens@hs-anhalt.de
Web: http://icaiit.org

Content

Section 1 Communication and Data Transport Technologies

Section 2 Data Analysis and Processing

Inverse and Direct Maxflow Problem Study on the Free-Oriented ST-Planar Network Graph

Victor Tikhonov[1], Serhii Nesterenko[2], Abdullah Taher[3], Olena Tykhonova[1], Olexandra Tsyra[1], Olha Yavorska[1] and Kateryna Shulakova[1,4]

[1]*Department of Computer Engineering and Information Systems, State University of Intelligent Technologies and Telecommunications, Kuznechna Str. 1, Odesa, Ukraine*

[2]*Department of Computer Intellectual Systems and Networks, Odesa Polytechnic National University, Shevchenko Avenue 1, Odesa, Ukraine*

[3]*Department of Electronic and Communication Engineering, College of Engineering, Al-Qadisiyah University, Iraq*

[4]*Anhalt University of Applied Sciences, Bernburger Str. 55, Köthen, Germany*

{victor.tykhonov, elena.tykhonova}@suitt.edu.ua, sa_nesterenko@ukr.net, abdallahqays@gmail.com, aleksandra.tsyra@gmail.com, yavorskayao7@gmail.com, katejojo29@gmail.com

Keywords: Telecommunication Network, Maximal Flow, Free Oriented Planar Graph, SDN.

Abstract: The issues of data flow optimization in telecommunication networks are considered. The analyses of the problem state of art shows the primarily utilization of logistic Maxflow model on ST-planar directed network graph with predetermined fixed metric. Concluded, that conventional logistic Maxflow model is not adequate to modern telecoms with flexibly reconfigured channels. Introduced the concept of the free-oriented network graph as an enhanced math-model for digital flows simulation. The inverse and direct Maxflow tasks are formulated on the normalised free-oriented ST-planar network graph, and the properties of the graph obtained as functions of vertices number. The direct Maxflow task is studied in tensor form, and the algorithm of test-sequences generation for the inverse Maxflow task is constructed. The inverse Maxflow problem has been analyzed as a discrete optimization task on the Pontryagin maximum principle with two necessary extremum conditions. Related computation algorithm is built with polynomial complexity. Unlike the known approaches, proposed method is relevant to data flow optimization in the software defined networks with dynamically reconfigurable channels. Along with the maximal flow, the flow distribution over the network structure provided. The formalism of the direct Maxflow task can be used for testing the algorithms of inverse Maxflow task solutions, and generation the training sequences for machine learning in AI models.

1 INTRODUCTION

An efficient way to reduce the cost and improve the quality of telecommunication service is increasing the network productivity by optimal scheduling the digital flows to gain better equipment utilization. This task is known as the Maxflow optimization problem. Historically, the Maxflow problem (MFP) arose in logistic systems to deliver some unimodal product (e.g. gas, oil, water etc) from a producer site S to consumer target T. A feature of such systems is that products should not simultaneously be sent in opposite directions, and transportation channels from source S to target T should not intersect with each other; besides, there should be no internal flow generators or accumulators other than S and T.

The Maxflow problem conventionally involves searching a feasible flow on the given ST-directed weighted planar network graph with single-product flows and two unique vertices – the flow generating source S and flow accumulating sink T. Such a graph presumes that all the graph-arcs have fixed weights. Denote this type of graph ST-DWPG.

The ST-DWPG graph as a generic logistic-system presentation with fixed arcs-weights still remains to be a common model for digital flows optimization. However, it is not fully adequate to modern telecommunication data networks with dynamic reconfiguration of channel capacities, potentially enabling the overall network performance increase.

This work intends to advance the Maxflow study in data networks, using the feature of dynamic reconfiguration in digital communication channels.

Section 2 of this work considers the Maxflow problem state of the art. Section 3 formulates the objectives of the work. Section 4 introduces a dual formalization of the Maxflow problem as traditional (inverse) task of feasible flow finding and a coupled (direct) task of generating the testing sequences for MFP solution algorithms. Section 5 gives definition of the direct Maxflow task in tensor form. Section 6 provides a discrete analysis of the inverse Maxflow problem on the ST-planar free-oriented graph. Section 7 summarizes results of the work.

2 THE MAXFLOW PROBLEM STATE OF THE ART

The widely known solution of the Maxflow problem (MFP) is Ford-Fulkerson algorithm (FFA), also referred to as "method" [1] (1956). It uses the Depth-First path Searching technique (DFS), and works by iteratively finding an augmenting ST-path in residual graph (obtained by subtracting the current flow from all the arc-capacities along the flow-path). The optimum-criterion at any iteration is "maximal flow-increment along the path".

The FFA stops when no more paths exist. It has a pseudo-polynomial run-time complexity O(F*E), where F is the maximal ST-flow, E is the number of arcs in the ST-DWPG graph.

Based on FFA, alternative approaches have been developed. Among them, a very popular and strongly polynomial algorithm for Maxflow problem solution was proposed by E. Dinitz [2] (1970). In contrast to FFA, it provides the vertex-based Breadth-First path Searching (BFS) technique to find a single source shortest path in an unweighted graph. It has the $O(V^2*E)$ computational complexity; V, E – vertices and edges numbers.

Similar to Dinitz, another FFS-based MFP-algorithm was published by J. Edmonds and R. Karp in 1972 [3]; it has $O(V*E^2)$ computational complexity. For certain network topology it outperforms the Dinitz method in computation time. The theoretical foundations of the Maxflow problem in the context of general graph theory applications have been outlined in [4] (1976).

A new push–relabel Goldberg-Tarjan algorithm for MFP solution proves that Dinitz blocking-flow method needs not more than $O(\min\{V^{2/3}, E^{1/2}\}$ augmentations in the unit capacity case [5] (1986). In particular case of capacity scaling and ingenious data structures, the run-time overcomes the O(V*E) barrier in Goldberg-Rao algorithm [6] (1998).

In 2006, an improved Goldberg-Rao algorithm for the maximum flow problem was exhibited by D. Papp [7]. It works not only for unit-capacity network graphs (as Dinitz algorithm does), but for more general case of binary weighted arks.

An original algorithm to approach the feasible flow in near linear time was proposed by P. Christiano et al for weighted undirected ST-graph. It uses the well-known Laplacian matrix method to solve a system of linear equations for electrical flow [8] (2010). The work claims developing the fastest known algorithm for computing approximately maximum S-T flows.

An enhanced formalism for combinatorial optimization problems like MFP is presented by M.A. Rajouh based on Pontryagin maximum principle. It formulates a set of necessary extremum-conditions kept along the algorithm iteration, that enables to curtail the initial diversity of iterations to a compact set of candidates for optimum. In some cases, the necessary conditions prove sufficient [9] (2013).

A solid overview of network flow algorithms on the closed bipolar ST-directed weighted graph reflects the book [10] (2019) by D.P. Williamson. The book extends the scope of network flow study for multicommodity task, and gives the detailed analysis of particular two-commodity case. Also, the notion of ST-flow in undirected graph is defined, where any edge has an arbitrary positive orientation and consist of two opposite-directed arcs of equal capacities.

Generalized MFP-statement to simulate the information flows in software defined network architecture is introduced in [11] (2019). In contrast to conventional logistic-network model in the form of bipolar closed ST-directed weighted planar graph (ST-DWPG), it defines the overall network capacity and its maximal flow on the open 3-pole non-planar free-oriented network graph. Related algorithm for the 6-vertex graph is presented in [12] (2019).

An important class of electric-power flows optimization (the Equal Maximum Flow Problem – EMFP) has been studied in [13] (2020). The EMFP aims for a maximum equal flow on all edges in a subset of the whole network graph edge-set. The paper extends the EMFP to Almost Equal Maximum Flow Problems (AEMFP) where electric-flow values differ according to a certain deviation-function.

A fast and near optimal algorithm for optimization of the end to end routing in SDN-network is proposed in [14] (2021), based on the max-flow/min-cut theorem. It shows that the maximal flow searching technique with minimal hops paths provides better performance compared to the traditional shortest path algorithms.

To optimize the complex multimodal flows in logistic systems, an application of genetic algorithm described in [15] (2022), which is generally used to solve the Vehicle Routing Problem (VRP). It simulates Darwin's natural evolution in transportation network model, and includes five recycling steps (task coding, population initialization, selective operations definition, mutation, selection process).

The overview of deterministic Maxflow and Min-cost flow logistic optimization problems for ST-directed weighted graph is given in [16] (2023) with emphasis on the worst-case of running time. The paper solely reflects the algorithms having a faster run-time than earlier published. It noticed, that Goldberg–Rao algorithm has a weakly polynomial run-time; the existence of a strongly polynomial O(V*E) algorithm was shown by Orlin in 2013, Orlin and Gong in 2021. Besides, interior-point methods (IPMs) for MFP-solutions considered, as well as the algorithms for minimal flow-cost.

A promising direction in solving complex combinatorial problems such as MFP is the use of artificial intelligence (AI) and machine learning (ML) technologies. A survey of the recent attempts in this area published in [17] (2021). Due to the hard nature of these problems, the surveyed algorithms rely on the handcrafted heuristics for making decisions, that otherwise are too expensive to compute or mathematically not well defined. Thus, machine learning looks like a natural candidate to make such decisions in a more principled and optimized way.

A set of single-machine scheduling problems with resource-dependent processing times is studied in [18] (2021). Heuristic algorithms for solving the reduced to unified model problems are presented. Three types of scheduling tasks are specified: due dates to jobs assignment, resources allocation to job operations and machines jobs scheduling.

Various researches on graph-based tasks deep learning are surveyed in [19] (2022). The paper focuses directed/undirected wired and wireless SDN networks as the most promising solution driving the networking industry to re-examine traditional network architecture. To track the follow-up research, a public GitHub repository is created, where the relevant papers will be updated continuously.

The ML-based approach to MFP solution presents the N. Orkun Baycik publication, where the tree-based learning method is applied. It includes decision tree and random forest regression (supervised learning algorithm and bagging technique, that uses an ensemble learning method). Both trees are built independently [20] (2022).

A network based intelligent node labelling (INL) algorithm for solving the Maxflow problem in directed network graph is developed in [21] (2023). It eliminates common augmenting paths technique to compute the maximal flow. Instead, it tries to balance input-to-output flow values for all the intermediate nodes, thus avoiding the excess or stagnant flow and reduction of the under-utilized outflow arcs. The algorithm needs at most two iterations to transform the initial N-nodes network into an equivalent network with O(V*E) worst-case complexity.

Summarizing the spoken above methods, algorithms and approaches to Maxflow problem solution, we conclude the following:

1) The most of the surveyed works on the Maxflow problem explore the so called "logistic model" of transportation system in the form of directed bipolar ST-graph with a source node S and a target node T [1-7, 10, 14-16, 19, 21]; also, undirected graphs are used to simulate electrical circuits and networks ([8, 13]). The case of 'multi-source/multi-target' graph (aka bipartite graph) can be easily reduced to a bipolar ST-graph. The case of undirected graph is commonly understood as a symmetrically directed graph.
2) The conventional Maxflow model of logistic system is not really adequate to modern data networks with dynamic reconfiguration of optical and wireless communication channels, and therefore, prevents to benefit the digital channel reconfigurability for data network performance. A further development of logistic flow model towards the SDN networking architecture is given in [11-12] by the concept of the 'free-oriented weighted graph' along with related algorithm of the Maxflow computation over the 3-pole/6-node open graph.
3) Recently, artificial intelligence (AI) and machine learning (ML) techniques have been increasingly used to solve complex combinatorial problems on graphs [17-21]. An important inductive AI-method is supervised learning (input objects and desired output-values for machine learning an AI-model). In terms of MFP, a graph sample G is an input, and the maximal flow distribution F on the graph G is the output. So, the sequence {G, F} can be used for ML. Yet, finding F for G is a hard problem; instead, constructing the graph G(F) for a given F is a rather trivial task. In all, generating a complete set {G(F)} is equivalent to building a sufficient ML-sequence {G, F}.

Based on the surveyed publications analysis, the objectives of this work are formulated further on with emphasis on data flow study in bi-polar networks.

3 OBJECTIVES OF THE WORK

This work aims to take use of the dynamic reconfiguration ability of the modern telecommunication channels for data networks performance increase. To achieve this, the following objectives have been set:

1) Formalization of the inverse and direct Maxflow problem on the free-oriented bi-polar ST-planar network graph with multiple vertices. This includes normalization of the complete ST-planar graph in visual and matrix forms with the distinguished topology and metric, studying the principal properties of the normalized graph, setting the direct and inverse Maxflow tasks.
2) Definition of the direct Maxflow task in tensor form. The set of ST-paths in the graph is given as a system of vectors (tensor) in the N-dimensional orthonormal Euclidean space, where N is the full number of edges in ST-planar graph.
3) Discrete analysis of the inverse Maxflow task on the free-oriented ST-planar network graph as an NPC-combinatorial optimization task, aimed to find the feasible flow between the two open graph poles, along with the flow distribution tensor.

4 FORMALIZATION OF THE INVERSE AND DIRECT MAXFLOW PROBLEM ON THE FREE-ORIENTED ST-PLANAR NETWORK GRAPH

The concept of the free-oriented weighted graph (FWG) with reconfigurable edges was exhibited in 2019 to apply the known MFP-algorithms to digital flows simulation in telecoms [11]. In contrast to other methods, the weight of any FWG-graph edge is assumed to be equal the fixed capacity of a duplex digital channel, that is divided in any proportion between the two coupled simplex-channels, on order to fully utilize the overall channel capacity.

Besides, the solution of the FWG-based Maxflow task does not only involve finding the feasible flow F_{max}, but also the distribution of this flow $D(F_{max})$ over ST-paths of the graph. Related algorithm for calculation F_{max} and $D(F_{max})$ on the 3-pole/6 vertex FWG-graph is presented in [12].

In this regard, there is a need in adaptation of the conventional ST-DWPG model of a logistic system for digital flows simulation in data networks with flexible configurable channels.

In graph theory, graph G is a cortege of binary relations g(k, m) on the set of vertices V: G(V):= {g(k, m)}, k, m∈V. If g(k, m) ≡ g(m, k), then G is undirected graph; else G is directed graph [22]. That means: a) G is directed graph, if at least one g(k, m) ≠ g(m, k); b) undirected graph is equivalent to symmetrically directed graph; c) the Ford-Fulkerson algorithm FFA is valid for both directed and undirected planar graphs.

Consider conventional ST-planar directed weighted graph G(V) of 4 vertices {S, T, 2, 3} and 6 weighted arcs: $g_{S,2} = g_{3,T} = 1$; $g_{S,3} = g_{2,T} = 3$; $g_{2,3} = g_{3,2} = 1$. It is easy to see, that here $F_{max} = 3$, wherein residual graph RG has 3 arcs: $g_{S,3} = g_{2,T} = g_{2,3} = 1$; the arc $g_{2,3}$ can't be used to augment the flow.

Instead, the equivalent free-oriented weighted model FWG(V) of the same case has 5 edges $g_{S,2} = g_{T,3} = 1$; $g_{S,3} = g_{T,2} = 3$; $g_{2,3} = 2$. By this model, F_{max} is increased to 4, wherein residual graph RG falls down to null. This illustrates the privilege of coherent transition from the common logistics directed graphs to the free-oriented graph models in telecoms.

Definition of normalized ST-planar graph. Consider the visual and matrix forms of the ST-planar free-oriented weighted graph (FWG); the first one is less formal but intuitively clear; the latter is more convenient for digital processing. It is clear, that both forms of the graph are invariant towards the vertex's names. Let graph G(V) has digital vertex's indexes 0, 1, 2, ..., V-1 (V is the total vertex's quantity). Let indexes 0 and 1 reserved for S and T network nodes (we label them as S0 and T1 in visual form). In these terms, Figure 1 depicts normalized 6-vertex ST-planar visual graph; figure 2 shows this graph in matrix view; the graph edges have numeric names (1÷12).

The visual graph framework (Figure 1) contains the nested sub-graphs, regularly growing by adding new vertices 2, 3, ... to the initial primitive graph of two adjacent vertices (S0, T1) with two outer open edges. Vertices S0 and T1 simulate digital network border gateways for transfer the data flow F in both directions via the open edges. This framework has maximal number of edges (12) for a 6-vertices planar graph. The graph topology in Figure 1 is shown by edge-lines, while the metric is given by the edge values.

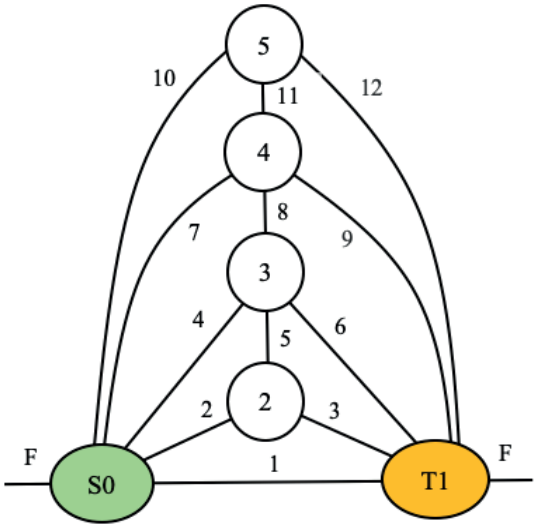


Figure 1: Normalized ST-planar visual graph.

The matrix graph (Figure 2) is divided in three sections: right-upper (topology), left-down (metric), diagonal (open edges). The yellow cells in Figure 2 indicate three eliminated edges of complete planar 6-vertex graph with 12 edges compare to complete non-planar graph with 15 edges. In digital presentation of matrix graph, names "S0", "T1" are to be substituted by 0 and 1 to adhere the Python array indexing [23].

	S0	T1	2	3	4	5
S0	1	1	1	1	1	1
T1	1	1	1	1	1	1
2	2	3		1	0	0
3	4	6	5		1	0
4	7	9	0	8		1
5	10	12	0	0	11	

Figure 2: Normalized ST-planar matrix graph.

In classic analysis, 'topology' T is a collection of subsets X` of the given set X, if: {∅, X∈T; ∀(U, V∈T) U∩V∈T; $\forall_I \cup(V_I \in T) \in T$}, where Ø is empty subset. The pair (X, T) is called 'topological space'. The subsets X`∈T are called 'open subsets' [24].

Considering the FWG-graph model, we interpret 'topology' as vertices connectivity in the free-oriented graph, where ∅ is open network environment; the 'open subsets' X` are bound sub-graphs that have finite paths to the network-outside via the open poles S and T; the latter simulate the network border gateways.

The visual graph in Figure 1 is convenient for studying core properties of ST-planar network graph as functions of the vertex quantity (V), which are important to formalize the inverse and direct MaxFlow problem in telecoms.

The number of inner edges. In case of non-planar complete graph, the number of inner edges equals $E_{NPG}(V)=V(V-1)/2$. Graph with V≤4 is always planar, so $E_{NPG}(3)=E_{PG}(3)=3$. In Figure 1, adding any next vertex to initial 3-vertex graph increments the total number of planar graph edges E_{PG} in 3 edges: $E_{PG}(3+1)=3+3=6$; $E_{PG}(4+1)=6+3=9$; $E_{PG}(5+1)=9+3=12$ and so on. In general, $E_{PG}(V)=3(V-2)$, V≥3.

The number of ST-paths. The ST-planar graph with 3 vertices (S0, T1, 2) in Figure 1 has two ST-paths: one 1-hop-path 'S0-T1' and one 2-hop-path 'S0-2-T1'. Each next vertex V adds two paths more: one 2-hop-path 'S0-V-T1' and one h-hop-path (h≥3) through the 'vertical' vertex (V-1, V), e.g. paths 'S0-3-T1' and 'S0-2-3-T1'.

Thus, the number of ST-paths in complete planar graph is P(V)=2(V-2), V≥3. This includes: one 1-hop-path, (V-2) of 2-hop-paths, (V-3) of h-hop-paths with h>2; in sum: 1+(V-2)+(V-3)=2(V-2), V≥3. Any next path p in the list of counted above paths P(6)=8 in ST-planar graph (Figure 1) has at least one edge, that is new -added to the previously listed edges (yellow marked cells in Figure 4), and in bold-font here: {p}=(**1**), (2, **3**), (**4,** 6), (7, **9**), (**10,** 12), (2, **5,** 6), (4, **8**, 9), (7, **11**, 12).

Distribution $P_V(h)$ of h-hop paths. We bring without proof the empirical formula: $P_V(h)=1$, if h=1; $P_V(h)=V-h$, if h=2, 3, ..., V-1; e.g. $P_6(h=1, 2, 3, 4, 5)=1, 4, 3, 2, 1$. The set of eight shortest paths includes: $P_6(h=1, 2, 3)$ = {1, 4, 3}. The set of 8 max-diverse and short paths is {$P_6(1)=1$; $P_6(2)=4$; $P_6(3)=1$; $P_6(4)=1$; $P_6(5)=1$} formed by edges {(**1**); (2, **3**); (**4**, 6); (**7**, 9); (10, **12**); (2, **5**, 6); (2, 5, **8**, 9); (2, 5, 8, **11**, 12)} with 8 backward-unique edges in bold. It can be argued, that the number of paths P(V)={p} equals the total number of additive flows f(p) in complete ST-planar graph G(V). Figure 3 summarizes the spoken above properties of ST-planar network graph.

V	3	4	5	6	Function
E_{PG}	3	6	9	12	3(V-2)
P(V)	2	4	6	8	2(V-2)
$P_V(h=1)$	1	1	1	1	1
$P_V(h=2)$	1	2	3	4	V-h
$P_V(h=3)$	0	1	2	3	V-h
$P_V(h=4)$	0	0	1	2	V-h
$P_V(h=5)$	0	0	0	1	V-h
$P_V(h<V)$					V-h

Figure 3: ST-planar free-oriented graph properties.

The known MFP-algorithms suppose predefined fixed arcs-weights, and thus, do not correctly work on the FWG-model with flexible scheduled arcs. We propose here an alternative approach to MFP study in the form of the direct and inverse Maxflow task.

The ordinal direct Maxflow task (O-DMT) we set as 'how to construct the free-oriented ST-planar graph G(F) with minimal edges` capacities to provide the given total flow distribution F over the set of ST-paths. The inverse Maxflow task (IMT) we set as 'how to find the maximal feasible ST-flow distribution F(G) over the set of ST-paths on the given free-oriented ST-planar graph G.

Generic inverse maxflow task IMT implies scanning all the ST-paths on the network graph topology. The total number of paths P(V) as function of vertices quantity V has a faster than polynomial growth even for planar graphs. Such computational tasks are referred to as nondeterministic polynomial complete problems (NPC) [13].

To solve complex combinatorial optimization problems of NPC-type, heuristic approaches are often used, which do not guarantee the best results, but may have less complexity, calculation time and produce adequate outcomes [18].

One of the famous and widely used heuristic method is Pontryagin maximum principle. This is choosing a set of necessary extreme conditions kept at any point of the object phase-track or the algorithm iteration, aimed to curtail the initial solutions diversity to a compact set of candidates for optimum. In some cases, the necessary conditions prove sufficient [9].

Solving complex NPC problem by heuristic method needs reliable proof the result obtained. The rigorous proof not always feasible, and empirical verification often necessary using various testing tools. Here, we propose an idea 'Verify complex inverse task algorithm by simple direct task samples generation'.

The spoken above ordinal direct Maxflow task O-DMT is rather simple for particular given flow F-distribution over ST-graph G. In general, the direct Maxflow task (DMT) turns into non-trivial case of 'generation a comprehensive sequence {F, G} with input F and output G for IMT-testing with inversed input/output {G, F}'.

Today, more and more human tasks addressed to artificial intelligence (AI) with digital neural networks, that capable to be taught by machine learning (ML) techniques. The matter is, how to construct training samples sequences (TSS) for AI-models comprehensive teaching.

Let S = (G, F) a distinct DMT-task output, which can be used for an IMT-case testing. In these terms, we define {S}= {(G, F)} as the testing/training samples sequence (TSS) for testing the inverse Maxflow task algorithms on the free-oriented ST-planar graph.

Towards the Maxflow problem in the context of AI, we propose the DMT-output generator of IMT-tests {G, F} to be used as a training samples sequence (TSS) for machine learning of an AI-model. The core issue of the direct Maxflow task (DMT) is, how many samples {S} needed for complete testing the IMT-algorithms or exhaustive MFP machine learning.

5 DEFINITION OF THE DIRECT MAXFLOW TASK IN TENSOR FORM

We formulate a particular case of direct Maxflow task (DMT) on the bi-pole planar FWG-graph as following: 'Find metric M of network graph G(V) with arbitrary number of vertices V, which is relevant to given flow distribution F(P) over the set P(V)={p} of paths $p \in P(V)$ in the given ST-planar free-oriented network graph G with complete topology T'. The 'complete topology' means, that all the possible edges E_{PG} of the ST-planar graph G(V) are included (Figure 3).

Consider the complete planar graph G(V) with V=6 vertices (Figure 1). The graph topology T is presented by the set of 12 edges $E_{PG}=\{e\}$ identified by their digital values 1÷12 in matrix graph (Figure 2). All the edges $e \in E_{PG}$ are mutually independent; let each of them be unitary (topological) vector **e**, and the set **E**:={**e**} be the Euclid vector basis. Graph G(6) has 8 paths P(6)=8, and each path $p \in P(V)$ possess at least one backward-unique edge e_f to carry augment flow f(p). The set of 8 shortest paths is a system of vectors (tensor) **P**={**p**} in matrix view (Figure 4); here, empty cells are zeros, **p**(1)=**e**(1); **p**(2)=**e**(2)+**e**(3) and so on. In this case, **P** includes: a single 1-hop path, four 2-hop paths and three 3-hop paths.

The matrix of scalar products {**p**(k)×**p**(m)} is *paths-metric tensor* M_P of **P** (Figure 5). It is evident, that the first five vectors of **P** are mutually orthogonal, as well as vectors number 6 and 8. The system **P** with $M_P>0$ we call the *path's tensor*. It is easy to show, that M_P is positive matrix [25].

Let $\{f_P(k)\}$ be the distribution of the entire ST-flow F_Σ over the paths p(k) of the graph G(V),

Figure 4; F_Σ=sum{$f_P(k)$}; **f**:={$\sqrt{f_P(k)}$} the *flow row-vector* with scalar product (**f**×**f**)=F_Σ; **F**:=diag(**f**) the diagonal matrix denoted as *flow-tensor*.

p(k)	e(n)											
n / k	1	2	3	4	5	6	7	8	9	10	11	12
1	1											
2		1	1									
3				1		1						
4							1		1			
5										1		1
6		1			1	1						
7						1	1	1				
8							1				1	1

Figure 4: System P of path' vectors in ST-planar graph.

The metric tensor M_F=(**F**×**F**) is distribution of the total flow F_Σ over the paths **P**={**p**}; F_Σ = trace(M_F)= (**f**×**f**). On this premise, we define the *flow-paths tensor* $\mathbf{F_P}$:=**F**×**P**. For the graph in Figure 1, the $\mathbf{F_P}$ is (8×12) matrix, where each k-row element is multiplied by the correspondent $\sqrt{f_P(k)}$ value.

n / k	1	2	3	4	5	6	7	8
1	1	0	0	0	0	0	0	0
2	0	2	0	0	0	1	0	0
3	0	0	2	0	0	1	1	0
4	0	0	0	2	0	0	1	0
5	0	0	0	0	2	0	0	1
6	0	1	1	0	0	3	1	0
7	0	0	1	1	0	1	3	1
8	0	0	0	0	1	0	1	3

Figure 5: Metric tensor M_P of path's P vector system.

The convolution g:=**f**×$\mathbf{F_P}$ we define as network *flow-load metric*. For the graph G in Figure 1, g is the row-set with 12 flow-loads of G-edges. If $f_P(k)\equiv 1$ and **P** in Figure 4, then **g**=(1, 2, 1, 1, 1, 3, 3, 1, 1, 1, 1, 2).

Now, the direct Maxflow task for ST-planar free-oriented graph G with complete topology T lies in generating a relevant manifold of path's tensors **P** on T, along with the metric tensors M_F. Here, the shortest-paths criteria or the maximal paths-diversity can be applied. This results in calculation the tensor $\mathbf{F_P}$=**F**×**P**=diag(**f**)×**P** and row-set g=**f**×$\mathbf{F_P}$:

$$\text{DMT: } (\mathbf{P}, M_F)_T \rightarrow (g)_T.$$

The graph G(T, g) with topology T and metric g is the input for the inverse Maxflow task, while the output is (M_F, **P**). Thereby, a comprehensive set {G(T,g)/(M_F,**P**)}, obtained by the routine direct Maxflow task solution, can be used for testing the non-trivial algorithms of the hard inverse Maxflow task, as well as a training sample sequence (TSS) for machine learning of an AI-model:

$$\text{IMT: } (g)_T \rightarrow (\mathbf{P}, M_F)_T.$$

Ultimately, the direct Maxflow task (DMT) as a first part of the whole Maxflow problem (MFP) for data flows simulation in telecommunication networks can be given by the following formalism.

1) Bring the network topology to the normalized view of the ST-planar graph G(V) with V vertices (see Figures 1 and 2).
2) Generate the path's tensor **P**(G) and flow-tensor **F** = diag(**f**), **f** = $\sqrt{f_P(k)}$ (see Figure 4).
3) Calculate the flow metric-tensor M_F = (**F**×**F**).
4) Get flow-paths tensor $\mathbf{F_P}$ = **F**×**P**(G).
5) Count the graph G(V) metric g = **f**×$\mathbf{F_P}$.
6) Fix (**P**, $M_F)_T \rightarrow (g)_T$ as a DMT result sample.
7) Use {$(g)_T \rightarrow$ (**P**, $M_F)_T$} as IMT/TSS sequence.

6 DISCRETE ANALYSIS OF THE INVERSE MAXFLOW TASK ON THE FREE-ORIENTED ST-PLANAR NETWORK GRAPH

Following the definitions in Section 5, here we introduce the inverse Maxflow task (IMT) on the free-oriented ST-planar weighted network graph G(V) with V vertices, complete topology T and given metric **g**, as the objective: "Find the feasible total flow value F_Σ between the two open poles S, T, along with the path's tensor **P**(G) and flow-path tensor $\mathbf{F_P}$".

This type of discrete optimization tasks belongs to the known class of NPC combinatorial problems. An obvious but worst strategy for exact solving the formulated above IMT task is as follows.

Let P_i={p_k}$_i$ the set of ST-paths on the complete ST-planar graph G, where each P_i is unique in paths p_k ordering, e.g.,

$$\{P_i\}=\{\{p_1, p_2, p_3\}, \{p_1, p_3, p_2\}, \{p_2, p_1, p_3\}, \{p_2, p_3, p_1\}, \{p_3, p_1, p_2\}, \{p_3, p_2, p_1\}\}.$$

On any iteration P_i, the path p_k is consequently examined for feasible flow F_i=sum{f(p_k)}$_i$ until no more paths p_k exist. Among the iterations {F_i}, one or more best results {F_{max}}⊂{F_i} are to be taken as final product. The power of the set {F_i} steamily grows with the increase of vertex's quantity V.

To reduce this task, apply the known heuristic approach on the base of Pontryagin maximum

principle. Let two extremum necessary conditions for each step ‘k’ of the current iteration ‘i’:

1) “max-flow-try”, i.e. f(pk)=max;
2) “shortest-path-first”, i.e. pk≤ pk+1 in hops.

The first condition is used in common FFA-based algorithms of MFP solution on the ST-planar directed graphs; this ensures convergence of the solution process.

The second condition is used in FFA-based Edmonds-Karp algorithm of MFP to reduce computational complexity. Figure 6 helps to see the necessity of the ‘shortest-path-first' condition to achieve the maximum flow in ST-planar free-oriented graph.

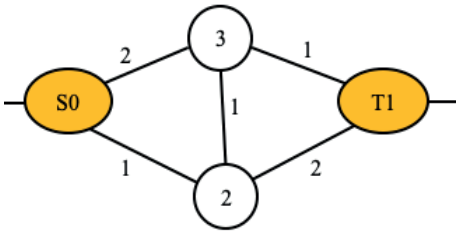


Figure 6: ST-planar graph case with critical first path.

It is clear, that the feasible flow from the source S0 to the target T1 in Figure 6 is $F_{max}= f_1$(S0-2-T1) + f_2(S0-3-T1) + f_3(S0-3-2-T1) = (1+1+1) = 3, whereas the residual graph is null. However, in case of violation the second necessary condition (‘shortest-path-first’), e.g. if begin with the 3-hop path p_1=(S0-2-3-T1) instead 2-hop path (S0-2-T1), we obtain less feasible flow F=1< F_{max}=3, and non-null residual graph G_R, whereas no more paths from S0 to T1 exist:

$$F=f1(S0\text{-}2\text{-}3\text{-}T1)=1<F_{max}=3;$$

$$GR=\{(3, S0)=2; (3, T1)= 2\};$$

vertices S0 and T1 are isolated.

Thus, violating the second necessary condition (‘shortest-path-first’) may result in false output product (at least one case shown above); therefore, it is always needed, i.e. is necessary condition.

Let estimate the complexity reduction of the inverse MFP-task by the ‘shortest-path-first’ condition provision. The total number of different h-hop paths in ST-planar graph G(V) with V vertices can be calculated with Figure 3:

$$NPV=1+sum(V-h),\ h\in 2, 3,...,V-1.$$

In particular case of V=6, it is

$$NP6=1+(6-2)+(6-3)+(6-4)+(6-5)=1=11.$$

It is easy to show, that in general case, the number of h-hop paths in ST-graph with V vertices is polynomial function

$$N_{PV}= 1+sum(V-2,V-3, ...,1) = 1+0.5(V-2)\ (V-1).$$

So, each set $P_i=\{p_k\}_i$ has N_{PV} paths number.

Let $\{P_i\}$ be the class of all the sets P_i. According to the known combinatorial formulas, the number of sets P_i in the class $\{P_i\}$ equals the number of permutations of the paths p_k:

$$|\{Pi\}|=1\cdot 2\cdot 3\cdot \ldots \cdot NPV=(NPV)!.$$

For instance, if V=6, then $|\{P_i\}|$=11!=39`916`800.

It is rather clear, that among all the P_i sets within the full $\{P_i\}$ collection, the only one set $P\in\{P_i\}$ may satisfy the "shortest-path-first" condition.

Thus, by applying Pontryagin maximum principle along with two heuristic necessary conditions (“max-flow-try” and “shortest-path-first”), the inverse combinatorial Maxflow task (IMT) on the bi-polar free-oriented ST-planar network graph can be reduced from the none-deterministic polynomial complexity NPC, with $\{(N_{PV})!\}$ iterations for scanning sets of paths $\{P_i\}$, to PC-task with computation the flow distribution over a single set of paths $P=\{p_k\}$, that includes the polynomial number of paths $N_{PV}= 1+0.5(V-2)\ (V-1)$.

Now, we get the answer, what ML training sequence TSS is sufficient for testing the IMT-algorithms on the FWG network graph. It is the set of paths N_{PV}, constructed by Pontryagin maximum principle with two extremum necessary conditions.

7 CONCLUSIONS

The main scientific result of the work is formalization of the inverse and direct Maxflow tasks on the free-oriented weighted ST-planar network graph. This allows to use the dynamic reconfiguration of the modern telecommunication channels in digital network-flow optimization, in order to increase the overall networks performance. Within the scope of this result, the following is obtained.

The Maxflow problem state of the art is analyzed in Section 2. It shows, that most publications on the Maxflow problem explore the so called “logistic models” of transportation system in the form of directed bipolar ST-graph. Among them, recent researches on artificial intelligence and machine learning techniques have been increasingly used to approach the Maxflow problem.

It is concluded, that conventional logistic Maxflow models are not enough adequate to modern

telecoms, and therefore, hamper to fully benefit the SDN-reconfigurability. Known algorithms on multi-pole free-oriented graphs are solely limited by the 6-node graph case. Because of that, further researches on Maxflow methods in telecoms on the free-oriented graph model needed.

To advance the Maxflow problem study in data networks with reconfigurable channels, related objectives have been set in Section 3. Section 4 formalizes the inverse and direct Maxflow tasks on the free-oriented bi-polar ST-planar network graph in terms of inverse Maxflow task testing and machine learning of artificial intelligence models. Section 5 studies the direct Maxflow task as a first part of the whole Maxflow problem in tensor form; an algorithm of testing samples set calculation is constructed.

The inverse Maxflow problem has been analyzed in Section 6 as a discrete optimization task on the Pontryagin maximum principle with two necessary extremum conditions: ‘max-flow-try’ for the flow-path scanning, and “shortest-path-first” for augmenting paths searching. The related algorithm is reduced to a single iteration of paths tensor analysis with polynomial paths number.

In general, unlike the known approaches to product flow maximization on logistic system model, a novel method introduced for digital flow optimization in software defined networks with dynamically reconfigurable channels. Along with the total maximal flow, this method provides the maximal flow distribution over the network structure. The direct Maxflow formalism also enables the Maxflow algorithms testing and machine learning of artificial intelligence models.

REFERENCES

[1] L.R. Ford and D.R. Fulkerson, "Maximal flow through a network," Canadian Journal of Mathematics, vol. 8, 1956, pp. 399-404. [Online]. Available: https://www.semanticscholar.org/paper/Maximal-Flow-Through-a-Network-Ford-Fulkerson/794b01b2513f3610cb151ecfd07e9dc0eae7e1f3.

[2] Y. Dinitz, "Algorithm for solution of a problem of maximum flow in a network with power estimation," 1970. [Online]. Available: https://www.researchgate.net/publication/228057696_Algorithm_for_Solution_of_a_Problem_of_Maximum_Flow_in_Networks_with_Power_Estimation.

[3] J. Edmonds and R. Karp, "Theoretical Improvements in Algorithmic Efficiency for Network Flow Problems," Journal of the Association for Computing Machinery, vol. 19, no. 2, 1972, pp. 248-264.

[4] J.A. Bondy, et al., "Graph theory with applications," Macmillan Press Ltd, GB, 1976. [Online]. Available: https://www.iro.umontreal.ca/~hahn/ift3545/gtwa.pdf.

[5] A.V. Goldberg and R.E. Tarjan, "A new approach to the maximum flow problem," Proceedings of the eighteenth annual ACM symposium on Theory of computing – STOC '86, 1986, pp. 136-146.

[6] A. V. Goldberg and S. Rao, "Beyond the flow decomposition barrier," J. ACM, vol. 45, no. 5, 1998, pp. 783-797.

[7] D. Papp, "The Goldberg-Rao algorithm for the maximum flow problem," COS 528 class notes, 2006, 5 p. [Online]. Available: https://www.cs.princeton.edu/courses/archive/fall06/cos528/handouts/Goldberg-Rao.pdf.

[8] P. Christiano, et al., "Electrical Flows, Laplacian Systems, and Faster Approximation of Maximum Flow in Undirected Graphs," Computer Science, Data Structures and Algorithms, 2010. [Online]. Available: https://www.semanticscholar.org/reader/af1afe809c106ed2618fc2ae14b696f672fc4bc1.

[9] M.A. Rajouh, "Osobenosti primenenia princip maximuma Pontryagin," BNTU, 2013. [Online]. Available: https://rep.bntu.by/bitstream/handle/data/5527/Osobennosti_primenenie.pdf?sequence=1&isAllowed=y.

[10] D.P. Williamson, "Network flow algorithms," Cornell University, 2019. [Online]. Available: https://www.networkflowalgs.com/book.pdf.

[11] O.V. Tykhonova, et al., "The Max-Flow Problem Statement on the Three-Pole Open Network Graph," 2019 3rd International Conference on Advanced Information and Communications Technologies (AICT), Lviv, Ukraine, 2019, pp. 209-212, doi: 10.1109/AIACT.2019.8847745. [Online]. Available: https://ieeexplore.ieee.org/document/8847745/metrics#metrics.

[12] O.V. Tykhonova, "Conveyor-modular method of multimedia flows integration with delay control in packet based telecommunication network," PhD Dissertation, O.S.Popov ONAT, Odessa, 2019.

[13] R. Haese et al., "Algorithms and Complexity for the Almost Equal Maximum Flow Problem," Operations Research Proceedings 2019. Springer, Cham, 2020, pp. 323-329.

[14] A. Alzaben, et al., "End-to-End Routing in SDN Controllers Using Max-Flow Min-Cut Route Selection Algorithm," 23rd International Conference on Advanced Communication Technology (ICACT), 2021, pp. 461-467.

[15] H. Liu, et al., "Optimization of a logistics transportation network based on a genetic algorithm," Hindawi Mobile Information Systems, vol. 2022, 2022, pp. 167-176. [Online]. Available: https://doi.org/10.1155/2022/1271488.

[16] O. Cruz-Mejía and A.N. Letchford, "A survey on exact algorithms for the maximum flow and minimum-cost flow problems," Networks, vol. 82, no. 2, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/net.22169.

[17] Y. Bengio, et al., "Machine learning for combinatorial optimization: A methodological tour d'horizon," European Journal of Operational Research, vol. 290, no. 2, 2021, pp. 405-421.

[18] B. Mor, et al., "Heuristic algorithms for solving a set of NP-hard single-machine scheduling problems with resource-dependent processing times," Computers & Industrial Engineering, vol. 153, 2021.

[19] W. Jiang, "Graph-based Deep Learning for Communication Networks: A Survey," Computer Communications, vol. 185, 2022, pp. 40-54. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0140366421004874?via%3Dihub.

[20] N. Orkun Baycik, "Machine learning based approaches to solve the maximum flow network interdiction problem," Computers & Industrial Engineering, vol. 167, 2022.

[21] T. Tawanda, et al., "An intelligent node labelling maximum flow algorithm," Int J Syst Assur Eng Manag 14, pp. 1276-1284, 2023. [Online]. Available: https://doi.org/10.1007/s13198-023-01930-3.

[22] "Graph theory," Britannica, 2023. [Online]. Available: https://www.britannica.com/topic/graph-theory.

[23] L. Wagner, "Complete Python Tutorial for Absolute Beginners," 2023.

[24] A. Kuronya, "Introduction to topology," 2010, 102 p.

[25] Matrix calculator. [Online]. Available: https://matrixcalc.org.

Methods of Spline Functions in Solving Problems of Telecommunication and Information Technologies

Irina Strelkovskaya[1], Irina Solovskaya[2] and Juliya Strelkovska[3]
[1]*Department of Information Technology, International Humanitarian University, Fontanskaya road Str. 33, Odesa, Ukraine*
[2]*Department of Computer Science, International Humanitarian University, Fontanskaya road Str. 33, Odesa, Ukraine*
[3]*Worthing College, 1 Sanditon Way, Worthing, United Kingdom*
{i.strelkovskaya, i.solovskaya}@mgu.edu.ua, 4800632s@gmail.com

Keywords: Spline Approximation, Spline Extrapolation, Real Spline Functions, Complex Spline Functions, Cubic Spline Function, Quadratic Spline Function, Hermite Cubic Spline, Signal Recovery, Forecasting, Self-Similar Traffic, IoT Device Telemetry Traffic, QoS Characteristics, Local User Location, DDoS Cyber-Attacks, Parametric Splines, 3D Modelling.

Abstract: The solution to problems of telecommunications and information technologies using spline approximation and spline extrapolation based on real and complex spline functions is considered. The use of spline approximation for solving problems of signal recovery and self-similar traffic, processes of functioning of telecommunication nodes and networks is shown. It is proposed to use spline extrapolation based on various types of real spline functions to solve the problems of predicting the characteristics of self-similar traffic and maintaining QoS characteristics during its maintenance. It has been established that to predict the real-time telemetry traffic of IoT devices, it is advisable to use spline extrapolation based on the cubic Hermite spike, which ensures the required forecasting accuracy and prevents network overloads, especially under conditions of network load limit. To solve the problem of user positioning in the radio access area, the use of complex plane spline functions is considered. The use of the methods of real and complex spline functions allows for obtaining the results of improving the quality of service in a telecommunications network and ensuring the scalability of the obtained solutions. To identify and predict DDoS cyber-attacks, a spline extrapolation method is used. The use of parametric splines in the problems of information technology, namely, the construction of curves and surfaces in 3D modelling, is proposed.

1 INTRODUCTION

The rapid development of technology leads to changes in many technological solutions and the improvement of protocols and algorithms of systems and software. Such changes require reviewing existing solutions and searching for new methods of solving telecommunications and information technology problems [1-4]. The process of creating a theory of telecommunications and information technologies is not yet complete; however, engineering intuition is often ahead of science today and allows finding the necessary solutions to problems at the intersection of different sciences, thereby moving from purely technological solutions to solutions based on system theory [5-8].

The solution of many problems requires the search for effective tools for system analysis and synthesis of modern telecommunication and information systems based on new mathematical methods. A review of the known solutions suggests that they are based on several already-known mathematical methods, such as functional analysis, integral and differential calculus, group theory, operator theory, spectral analysis, and others [1-8]. It is important to find such mathematical methods, an alternative to the known ones, which will significantly simplify the solution process and at the same time provide new solution possibilities. Such a solution may be the use of spline function methods, such as spline approximation and spline extrapolation.

The studies carried out by the authors in [9-12], allow us to assert the advantage of using splines, because: splines are easy to calculate, have good convergence, are resistant to local disturbances, have scalability properties, and provide opportunities to ensure the required accuracy.

The use of interpolation, approximation, and extrapolation methods for processes and modes of network elements, as well as network functions, allows us to obtain solutions to a significant class of telecommunications problems using real and complex spline functions [9-14]. The use of real and complex spline functions has an important practical advantage - a fairly simple implementation and the possibility of almost universal use.

Paper [9] shows the recovery of signals using counts and compares the recovery results using the Kotelnikov series and spline approximation based on cubic splines. It is established that the use of spline approximation for a signal with frequency fluctuations allows for obtaining results with the smallest error.

The advantage of spline approximation in solving traffic restoration problems has been proven using various types of spline functions (linear, quadratic, cubic, B-splines, etc.) in [9-12].

At the present stage of telecommunications development, an important issue is to solve the problems of predicting the characteristics of traffic served in the network. For this purpose, works [10-11] propose the use of spline extrapolation based on various spline functions (linear, quadratic, quadratic B-splines, cubic, cubic B-splines and cubic Hermite splines), which allow to increase the accuracy of the forecast.

The forecasting results allow us to predict the required size of buffer devices and characteristics of network objects, thereby preventing network congestion and exceeding the standard values of QoS characteristics [11].

However, there are several problems that cannot be solved using real splines, such as positioning problems. When solving such problems, there is a need to determine the user's coordinates about the received signal strength.

In this case, the authors propose the use of complex plane spline functions, which allow for to reduction of the positioning error compared to other known methods [12].

Today, the IT industry needs a different approach to solving a significant number of tasks that require a reduction in computational complexity and calculation time, but have requirements for the accuracy of the results. For example, in 3D modelling, cybersecurity and information protection tasks [9-14].

The purpose of this paper is to find new methods for solving analysis and synthesis problems in information technology.

2 APPROXIMATION ON THE BASIS OF SPLINE FUNCTIONS IN PROBLEMS OF TELECOMMUNICATION AND INFORMATION TECHNOLOGIES

With the help of real spline functions (linear, quadratic, quadratic B-splines, cubic, cubic B-splines and cubic Hermite splines), several problems in telecommunications have been solved, namely:

- the restoration of random signals and self-similar traffic, the solutions of which allow obtaining the required values between interpolation nodes with the required error [9-10];
- management of network objects and the network as a whole, which are based on the results of network monitoring and data processing [11];
- support of the procedures for the operation of objects and the network as a whole to improve the quality characteristics of QoS/QoE operation [11].

2.1 The Problem of Restoring Continuous Signals Using Spline Approximation Based on Real Cubic Splines

Consider the problem of restoring a continuous signal $f(x)$ on the interval [0;1], which is given by a function of the form [9]:

$$f(x) = 2\frac{\sin(16\pi(x-\frac{1}{2}))}{16\pi(x-\frac{1}{2})} - 3\frac{\sin(8\pi(x-\frac{1}{2}))}{8\pi(x-\frac{1}{2})}, \quad (1)$$

using the Kotelnikov series and spline approximation. The original signal $f(x)$ is limited by the spectrum $F_{\max}$ = 8 kHz, the sampling step is. The recovery of the continuous signal $f(x)$ by the Kotelnikov series is shown in Figure 1.

Let us consider the recovery of the signal $f(x)$ using a spline approximation based on a real cubic spline. We use the samples of the original signal $f(x)$ as interpolation nodes.

Considering the signal $f(x)$ of the form (1) on the interval [0;1], we set the grid node partitioning Δ: $0 = x_0 < x_1 < ... < x_N = 1$.

To construct the cubic interpolation spline $S_3(x)$, we construct a cubic function on each of the intervals $[x_i, x_{i+1}]$, $i = 0, \ldots, n-1$.

The cubic spline $S_3(x)$ for $x \in [x_i, x_{i+1}]$, $i = 0, \ldots, n-1$ is as follows [9]:

$$S_3(x) = f_i(1-t)^2(1+2t) + f_{i+1}t^2(3-2t) + \\ + m_i h_i t(1-t)^2 - m_{i+1}h_i t^2(1-t), \quad (2)$$

moreover $t = (x - x_i)/h_i$, $S_3(x_i) = f_i$,

$S_3(x_{i+1}) = f_{i+1}$, $m_i = S'(f; x_i)$, $h_i = h = (b-a)/n$,

n – number of segment breakdown elements $[a,b]$.

The obtained recovery of the continuous signal $f(x)$ is shown in Figure 1. where line 1 is the original signal $f(x)$, line 2 is the value of the function $f(x)$ in the grid nodes, line 3 is the recovered continuous signal $f(x)$ using the Kotelnikov series, line 4 is the recovered signal $f(x)$ using the cubic spline approximation.

According to the results of the recovery of a continuous signal $f(x)$, it was found that the use of a spline approximation based on a cubic spline has a smaller error and can be used to recover various signals characterized by the presence of rapid oscillations [9].

Thus, it can be noted that in solving telecommunication problems, the use of spline approximation makes it easier to obtain solutions to a class of problems, such as restoring and estimating states (data, signals, traffic), signal and image processing tasks, including data filtering and compression, signal detection and measurement, and improving the QoS characteristics of telecommunication networks with the required accuracy.

2.2 The Problem of Spline Approximation of Curves and Surfaces Based on Real Parametric Linear Splines

Consider the task of creating curves and surfaces in 3D modelling. Spline curves and surfaces are used in animation, video games, and interactive applications. Creating the trajectory of objects in time, and deformation processes in response to user actions or changes in the environment is a complex and resource-intensive process. The mathematical apparatus of spline functions, which has been used in telecommunications applications, can help.

To solve modelling problems, namely, the construction of curves and surfaces, we use linear parametric splines.

When interpolating the curve given parametrically by (3), we divide the interval of change of the parameter u, thus $u_0 < u_1 < \ldots < u_N$.

Let us find the value of the function at the partition points u_i, $i = 0, N$ [15]:

$$\begin{cases} x_i = x(u_i); \\ y_i = y(u_i). \end{cases} \quad (3)$$

The interpolation parametric spline of the first degree on the interval between points P_i and P_{i+1} has the form [15]:

$$\begin{cases} S_1(x;s) = (1-t)x_i + tx_{i+1}; \\ S_1(y;s) = (1-t)y_i + ty_{i+1}. \end{cases} \quad (4)$$

where $t = (s - s_i)/l_i$, $l_i = s_{i+1} - s_i$, $i = 0,1,\ldots,N-1$.

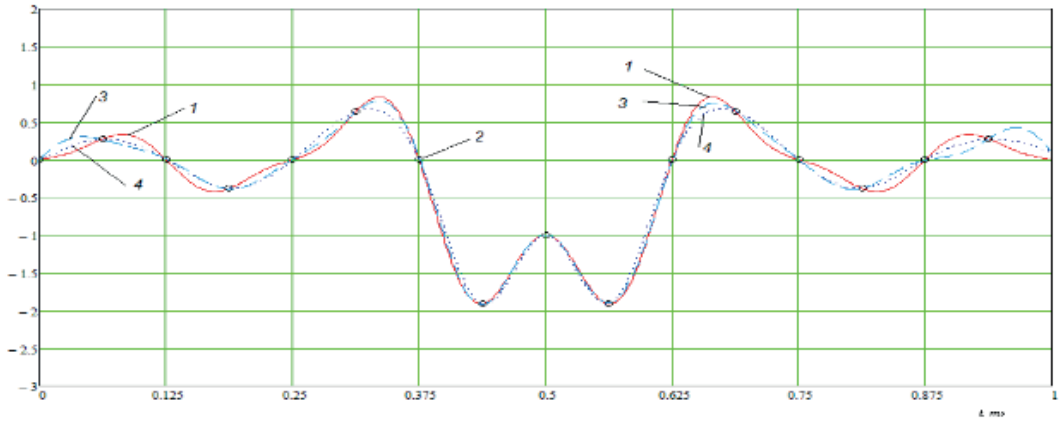


Figure 1: Restoration of a continuous signal $f(x)$ using the Kotelnikov series and spline approximation based on the cubic spline.

The set of splines $S_1(x;u)$ and $S_1(y;u)$ are a linear interpolation parametric spline. Depending on the type of function, linear parametric quadratic, cubic and other splines are considered [15].

An example of the construction of a linear parametric spline of the form (4) for a given curve L is shown by the dashed line in Figure 2.

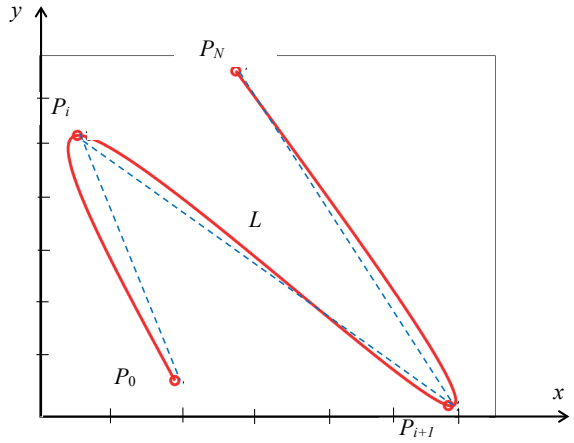


Figure 2: Approximation of a plane curve by a linear parametric spline.

It is easy to see that such a curve has significant errors of up to 25 %, which can be reduced by reducing the interpolation step.

To reduce the error in the reproduction of curves and surfaces, it is possible to use higher-order splines, such as quadratic, cubic, or cubic B-splines.

2.3 The Problem of Spline Approximation Based on Complex Plane Splines

To determine the location of a user in a Wi-Fi/Indoor network, we will consider spline approximation [12-13], [16-18]. We will use a complex plane quadratic spline as an approximating function [12], [19-20].

Consider a Wi-Fi/Indoor network (Figure 3), which consists of a set of AP$_i$, where i is the number of AP$_i$, $i=\overline{1,3}$.

Define the coverage area of the Wi-Fi/Indoor network by $\overline{G}$, for which $\overline{G}\subset Q$ (Figure 4), where $Q=[a,a+H]\times[b,b+H]$ with the side $H>0$ and a step $h_N=\frac{H}{N}$, N is a natural number, $x_k=a+kh_N$, $y_j=b+jh_N$, $k,j=0,\ 1,\ldots,\ N$.

Moreover, $\overline{G}=G\cup\partial G$ where ∂G is the boundary of the domain G. Let's break down the coverage area of the Wi-Fi/Indoor network into segments $Q_{k,j}$,.

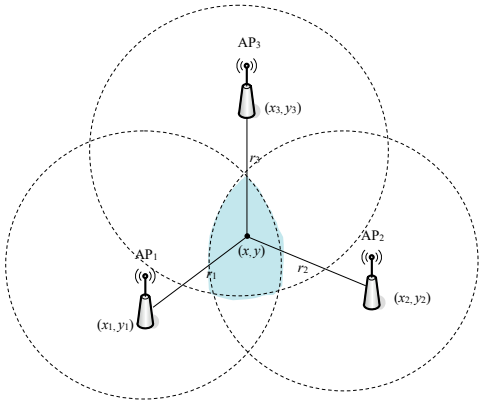


Figure 3: Wi-Fi/Indoor network.

Then $Q=\bigcup_{k,j=0}^{N-1}Q_{k,j}$,

$$Q_{k,j}=\left\{z=x+iy: x\in[x_k,x_{k+1}]\ ,y\in[y_j,x_{j+1}]\right\}.$$

This partition is denoted by Δ_N.

We define the domain G_N as the union of all segments $Q_{k,j}$ for which $Q_{k,j}\cap G\neq\varnothing$.

To find the coordinates of the user's location, consider one of the partitioning elements that are part of the G_N region with vertices P_1, P_2, P_3, and P_4, for which the following conditions are met (Figure 4) [12], [19-20]:

$$\begin{aligned}&\text{Re } P_1=\text{Re } P_4,\ \text{Re } P_2=\text{Re } P_3,\\ &\text{Im } P_1=\text{Im } P_2,\ \text{Im } P_3=\text{Im } P_4.\end{aligned}\qquad(5)$$

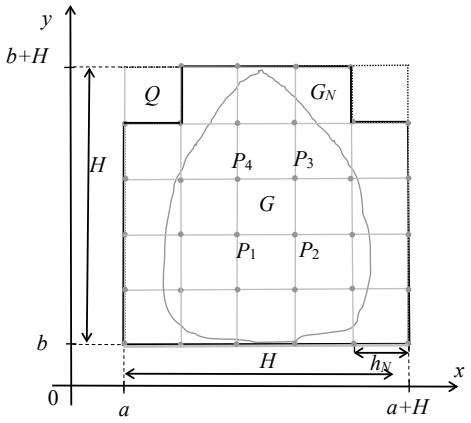


Figure 4: G_N area.

Let us construct a quadratic flat complex spline $S_\Delta(z)$, on the grid domain G_N that interpolates the function f(z) (or its continuation) at the vertices of the rectangles $Q_{k,j}$ included in the domain G_N, considering [12], [19-20]:

$$S_\Delta(z) = a + bz + c\bar{z} + d(z^2 - \bar{z}^2), \qquad (6)$$

at $z \in Q_{k,j} \subset G_N$, where the coefficients a, b, c, d are determined from the interpolation conditions at the $z_{k,j} = x_k + iy_j$:

$$S_\Delta(z_{k,j}) = f(z_{k,j}), \quad S_\Delta(z_{k+1,j}) = f(z_{k+1,j}),$$
$$S_\Delta(z_{k,j+1}) = f(z_{k,j+1}), \quad S_\Delta(z_{k+1,j+1}) = f(z_{k+1,j+1}).$$

The function $S_\Delta(z)$ is continuous in G_N [19-20].

3 EXTRAPOLATION ON THE BASIS OF SPLINE FUNCTIONS IN TELECOMMUNICATION AND INFORMATION TECHNOLOGIES

Spline extrapolation methods based on real spline functions of linear, quadratic, quadratic B-splines, cubic, cubic B-splines and cubic Hermite splines [15], [20] have allowed solving a significant class of problems in telecommunications and information technology:

- prediction of traffic characteristics, the solution of which in real time for different types of traffic (data traffic, voice traffic, telemetry traffic and video streaming traffic) provides opportunities to support QoS characteristics [11];
- support of QoS quality characteristics, namely, characteristics of delay time and probability of packet loss and distortion when servicing different types of traffic and forming requirements for network buffer devices [11];
- selection of the optimal configuration of telecommunication network objects, the decision of which is based on the results of forecasting the characteristics of the traffic of objects and the network as a whole to improve the quality of the telecommunication network [10-11];
- determination and forecasting of DDoS-type cyberattack traffic (SYN-Flood, ICMP-Flood, UDP-Flood) using splines [14].

3.1 The Task of Predicting the Characteristics of IoT Device Telemetry Traffic Using Spline Functions

When solving the problem of predicting the characteristics of the real-time telemetry traffic of an IoT (Internet of Things) device, it is taken into account that the traffic is created by sensors that operate according to a schedule and have functional features that determine the functioning of the system as a response to data requests to the sensors.

For forecasting, we use spline extrapolation based on the Hermite cubic spline [15], [20].

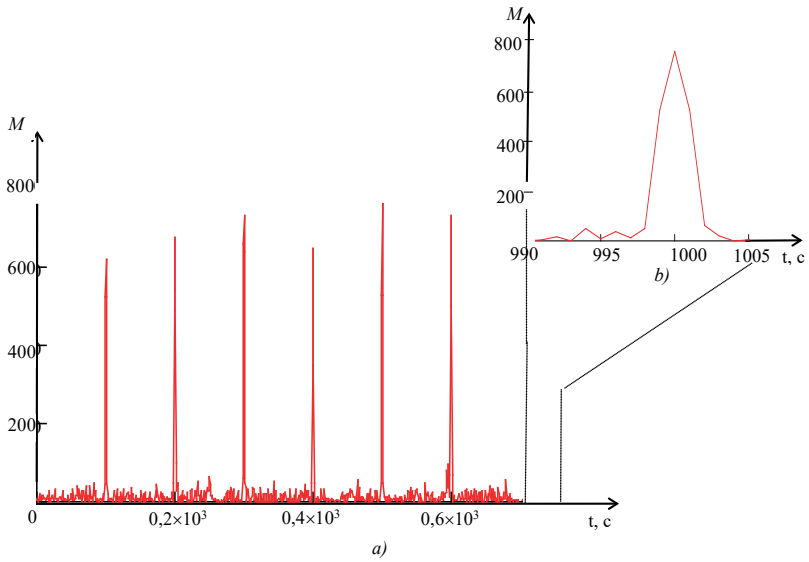


Figure 5: Real-time telemetry traffic of an IoT device, *a*) real traffic at [1;990] s, *b*) forecast traffic at [990;1005] s.

Consider the traffic of an IoT device on the segment $[a;b]$. To do this, let's break down $[a;b]$ as follows Δ: $a = x_0 < x_1 < \ldots < x_n = b$.

The nodes x_i contain the value of the function $f(x)$ and its derivative $f'(x)$, where $f_i = f(x_i)$, $f'_i = f'(x_i)$, $i = 0, 1, \ldots, n$.

Let's use the Hermite cubic spline $S_{3,2}(f;x) = S_{3,2}(x)$, that meets the conditions of [15], [20]:

- at each of the segments $[x_i; x_{i+1}]$, $i = 0,.., n\text{-}1$,

$$S_{3,2}(x) = a_{io} + a_{i1}(x\text{-}x_i) + a_{i2}(x\text{-}x_i)^2 + a_{i3}(x\text{-}x_i)^3;$$

- function $S_{3,2}(x_i) = f_i$, $S'_{3,2}(x_i) = f'_i$, $i = 0, \ldots., n$.

The Hermite cubic spline has the form [15], [20]:

$$S_{3,2}(x) = \varphi_1(t) f_i + \varphi_2(t) f_{i+1} + \varphi_3(t)\, hf'_i + \varphi_4(t)\, hf'_{i+1}, \quad (7)$$

moreover $\varphi_1(t) = (1 - t)^2 (1 + 2t)$, $\varphi_2(t) = t^2 (3 - 2t)$, $\varphi_3(t) = t (1 - t)^2$, $\varphi_4(t) = - t^2 (1 - t)$, $h_i = x_{i+1} - x_i$, $t = (x - x_i)/ h_i$, $i = 0, \ldots., n$.

The telemetry traffic of an IoT device is shown in Figure 5, where M is the number of packets (thousand packets).

The results of spline extrapolation of the traffic on the segment [990;1005] ms using the Hermite cubic spline $S_{3,2}(x)$ of the form (7) are shown in Figure 6.

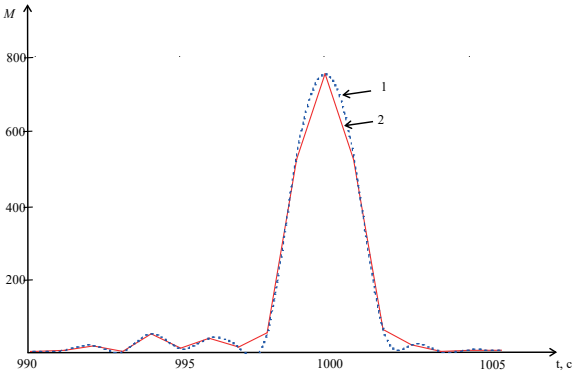


Figure 6: Predicting IoT device telemetry traffic using spline extrapolation based on the Hermite cubic spline: line 1 is real IoT device traffic, line 2 is predicted using spline extrapolation.

The obtained results of predicting the characteristics of the IoT device's real-time telemetry traffic allow for predicting the required IoT device bandwidth, as well as preventing congestion, which will ensure the maintenance of the required QoS characteristics.

3.2 The Task of Detecting and Predicting DDoS Cyberattack Traffic Using Spline Functions

When considering the characteristics of DDoS (Distributed Denial of Service) traffic, periodic or non-periodic, short and significant "bursts" of traffic intensity are often noted, while legitimate traffic has a small amplitude of pulsations that occur over a long period [21-30].

Given that cyberattack traffic has self-similarity properties, spline extrapolation can be used to solve the problems of detecting and predicting DDoS attacks (SYN-Flood, ICMP-Flood, UDP-Flood) [23-27]. We use the spline extrapolation method to predict the traffic of DDoS cyberattacks (Figure 7) [28-30].

To construct the cubic interpolation spline $S_3(x)$, we use the approach given in Section 2.1 and the expression of the cubic spline (2), for which we use the boundary conditions [15]:

$$S'(f;a) = f'(a),\ S'(f;b) = f'(b). \quad (8)$$

Considering the traffic on the segment $[a;b]$, we set a uniform partitioning grid with a step $h_i = h$, $i = 0, 1,\ldots, n - 1$, $h = \frac{b-a}{n}$. To build a cubic spline, we use the values of traffic intensities set at the interpolation nodes x_i, $i = 0, 1,\ldots, n$.

By constructing a cubic interpolation spline of the form (2) on the segment $[a;b]$, we obtain the required extrapolated values.

The proposed spline extrapolation based on spline functions has several advantages over the known methods. It is quite simple to implement, has a smaller error in predicting cyberattack traffic, and can be used to identify and predict DDoS traffic in real-time.

A successful choice of the type of spline functions when performing spline extrapolation can improve the accuracy of DDoS attack traffic detection. Prospects for further research are to further improve the accuracy of determining and predicting the traffic characteristics of various DDoS cyberattacks (SYN-Flood, ICMP-Flood, UDP-Flood) using the wavelet extrapolation method with an appropriate choice of wavelet functions [14].

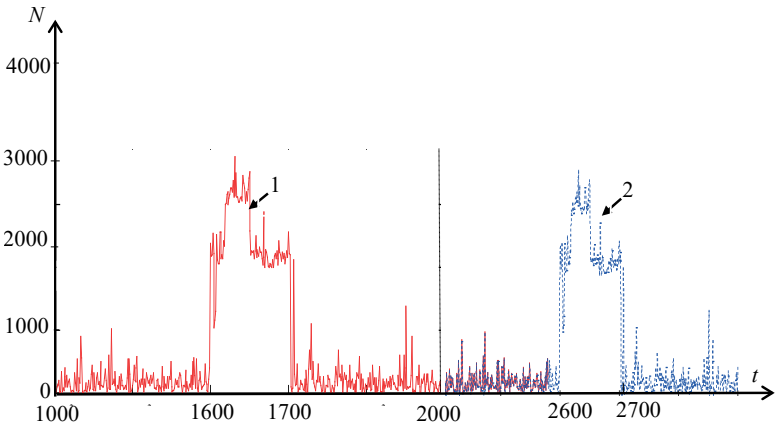


Figure 7: Results of extrapolation of DDoS cyberattack traffic on the segment [2600;2700] using a cubic spline: line 1 is DDoS attack traffic, line 2 traffic extrapolation using a cubic spline.

4 CONCLUSIONS

1) Along with solving telecommunications problems using spline approximation, spline approximation and spline extrapolation are proposed and used in cybersecurity problems, which can significantly simplify the solution process and at the same time provide new opportunities for obtaining results.

2) The use of spline approximation is proposed, which makes it easier to obtain solutions to such problems as restoration and estimation of states (data, signals and traffic), signal and image processing tasks, including the tasks of improving the quality characteristics of QoS of telecommunication networks.

3) A spline extrapolation method has been developed that allows us to obtain solutions to the problems of predicting traffic characteristics, identifying and predicting DDoS attacks, and improving the accuracy of the forecast, ensuring the scalability of solutions.

4) The obtained results of forecasting IoT device telemetry traffic allow to improve the accuracy of the forecast and ensure its scalability and use for various IoT applications, thereby avoiding network overloads and exceeding the normative values of QoS characteristics.

5) It is determined that for solving positioning problems it is advisable to use complex plane splines, which make it possible to simplify the determination of the user's location coordinates in the radio access network and increase the positioning accuracy.

6) The use of linear parametric splines is proposed for modelling curves and surfaces. It has been established that interpolation of curves and surfaces in 3D modelling using parametric splines has a higher accuracy of building curves and surfaces compared to other methods of restoring these curves and surfaces.

REFERENCES

[1] M. Ilchenko, L. Uryvsky, and A. Moshynska, "Developing of telecommunication strategies based on the scenarios of the information community," Cybern. Syst. Anal. 53, pp. 905-913, 2017. [Online]. Available: https://doi.org/10.1007/s10559-017-9992-9.

[2] M. Ilchenko, L. Uryvsky, and S. Osypchuk, "World trends of modern information and telecommunication technologies development," In: International Conference on Radio Electronics & Info Communications: Odesa, Ukraine, 2019.

[3] L. Globa, S. Dovgiy, O. Kopiika, and O. Kozlov, "Approach to uniform platform development for the ecology digital environment of Ukraine," In: Ilchenko, M., Uryvsky, L., Globa, L. (eds.) Progress in Advanced Information and Communication Technology and Systems. MciT 2021. Lecture Notes in Networks and Systems, vol. 548, pp. 83-100, 2023. [Online]. Available: https://doi.org/10.1007/978-3-031-16368-5_4.

[4] J. Donovan, "Building the Network of the Future: Getting Smarter, Faster, and More Flexible with a Software Centric Approach," Donovan John, Prabhu Krish (eds.). Chapman and Hall/CRC, 2017.

[5] O. Lemeshko, O. Yeremenko, and M. Yevdokymenko, "Tensor Model of Fault-Tolerant QoS Routing with Support of Bandwidth and Delay Protection," In: XIIIth International Scientific and Technical Conference Computer Sciences and Information Technologies, pp. 135-138, 2018.

[6] O. V. Lemeshko, O. S. Yeremenko, N. Tariki, and A.M. Hailan, "Fault-tolerance improvement for core and edge of IP network," 2016 XIth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 2016, pp. 161-164.

[7] O. Lemeshko, O. Yeremenko, M. Yevdokymenko, A. Shapovalova, A. M. Hailan, and A. Mersni, "Cyber Resilience Approach Based on Traffic Engineering Fast ReRoute with Policing," 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Metz, France, 2019, pp. 117-122, doi: 10.1109/IDAACS.2019.8924294.

[8] L. Uryvsky and K. Martynova, "Complex analytical model of priority requires service on cloud server," International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo-2019), 9-13 Sept. 2019, Odesa, Ukraine. [Online]. Available: https://doi.org/10.1109/UkrMiCo47782.2019.9165323.

[9] I.V. Strelkovskaya, E.V. Lysyuk and R.V. Zolotukhin, "Comparative analysis of restoration of continuous signals by Kotelnikov series and spline functions," East-European Journal of Advanced Technologies, vol. 2, no. 9(62), 2013, pp. 12-15.

[10] I. Strelkovskaya, I. Solovskaya, N. Severin, and S. Paskalenko, "Spline approximation based restoration for self-similar traffic," Eastern-European Journal of Enterprise Technologies, 2017, no. 3/4 (87), pp. 45-50.

[11] I. Strelkovskaya and I. Solovskaya, "Using spline-extrapolation in the research of self-similar traffic characteristics," Journal of Electrical Engineering, vol. 70, no. 4, 2019, pp. 310-316.

[12] I. Strelkovskaya, "Fingerprinting/Indoor positioning using complex planar splines," Journal of Electrical Engineering, vol. 72, 2021, no. 6, pp. 401-406. [Online]. Available: https://doi.org/10.2478/jee-2021-0057.

[13] V.I. Belyi and I.V. Strelkovskaya, "Approximation of functions by analytic complex splines in domains quasiconformal boundary," Ukrainian Mathematical Journal, vol. 40, no. 5, 1988, pp. 481-486.

[14] S. Kivalov and I. Strelkovskaya, "Detection and prediction of DDoS cyber attacks using spline functions," IEEE TCSET 2022: 16th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), Slavske, Ukraine, February 22 – 26, 2022.

[15] J.H. Ahlberg, E.N. Nilson, and J.L. Walsh, "The Theory of Splines and Their Applications," Academic Press, New York, 1967.

[16] S.H. Hameedah, M. Hussein, S. Mad Saad, and A. MatDzahir, "An overview of local positioning system: technologies, techniques and applications," Int. J. Eng. Technol., vol. 7, no. 3, 2018, pp. 1-5. [Online]. Available: https://doi.org/10.14419/ijet.v7i3.25.17459.

[17] F. Gu, et al., "Indoor localization improved by spatial contexts. A survey," ACM Comput. Surv., vol. 52, no. 3(64), 2019, pp. 1-35. [Online]. Available: https://doi.org/10.1145/3322241.

[18] Q. Yang, S. Zheng, M. Liu, et al., "Research on Wi-Fi indoor positioning in a smart exhibition hall based on received signal strength indication," Jur. Wirel. Commun. Netw., no. 275, 2019, [Online]. Available: https://doi.org/10.1186/s13638-019-1601-3.

[19] G. Opfer and M. Puri, "Complex planar splines," J. Approxim. Theory, vol. 31, no. 4, 1981, pp. 383-402.

[20] L.L. Schumaker, "Spline Functions: Basic Theory," Cambridge University Press, New York, 2007.

[21] D. Evans, "The Internet of Things. How the Next Evolution of the Internet Is Changing Everything," Cisco White Paper, April, 2011.

[22] J. Mocnej, A. Pekar, W. K.G. Seah, and I. Zolotova, "Network Traffic Characteristics of the IoT Application Use Cases," School of Engineering and Computer Science, Victoria University of Wellington, 2018.

[23] F. Tang, Z. Md. Fadullah, B. Mao, and N. Kato, "An Intelligent Traffic Load Prediction-Based Adaptive Channel Assignment Algoritm in SDN-IoT: A Deep Learning Approach," IEEE Internet of Things Journal, vol. 5, no. 6, 2018. [Online]. Available: https://doi.org/10.1109/JIOT.2018.2838574.

[24] M.F. Iqbal, M. Zahid, D. Habib, and L.K. John, "Efficient Prediction ob Network Traffic for Real-Time Applications," Journal of Computer Networks and Communications, vol. 1, 2019. [Online]. Available: https://doi.org/10.1155/2019/4067135.

[25] N.S. Khan, S. Ghani, S. Haider, and N.S. Khan, "Real-Time Analysis of a Sensor's Data for Automated Decision Making in an IoT-Based Smart Home," Sensors (Basel), vol. 18, no. 6, 2018. [Online]. Available: https://doi.org/10.3390/s18061711.

[26] M. Pivarníková, P. Sokol, and T. Bajtoš, "Early-Stage Detection of Cyber Attacks," Information, vol. 11, 2020, p. 560.

[27] D. Alghazzawi, O. Bamasag, H. Ullah, and M.Z. Asghar, "Efficient Detection of DDoS Attacks Using a Hybrid Deep Learning Model with Improved Feature Selection," Appl. Sci., vol. 11, 2021, p. 11634.

[28] T. Radivilova, L. Kirichenko, O. Lemeshko, et al., "Analysis of anomaly detection and identification methods in 5G traffic," 2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2021.

[29] B. Petrik and V.I. Dubrovin, "Detection of DoS attacks in network traffic by wavelet transform," Applied questions of mathematical modelling, vol. 4, no. 1, 2021, pp. 186-196. [Online]. Available: https://doi.org/10.32782/KNTU2618-0340/2021.4.1.20.

[30] P. Dymora and M. Mazurek, "Network Anomaly Detection Based on the Statistical Self-similarity Factor," Lecture Notes in Electrical Engineering, vol. 324(1), 2021, pp. 271-287. [Online]. Available: https://doi.org/10.1007/978-3-319-11248-0_21.

Overview of the Approaches to Managing Distributed Storage and Access to Cloud Data

Anton Kartashov and Larysa Globa
Institute of Telecommunication Systems, Igor Sikorsky Kyiv Polytechnic Institute, Beresteiskyi Avenue 37, Kyiv, Ukraine
kartashov.anton.ukr@gmail.com, lgloba@its.kpi.ua

Keywords: Cloud Storage, Cloud Access, Multi-Cloud.

Abstract: Currently, the main computing standard for hosting and delivering services over the global Internet has become the widely recognized Cloud Computing technology. Firms and end-users are constantly incorporating Cloud platforms into their technology stack because of their many advantages over traditional computing models. Key advantages of the cloud include nearly unlimited data storage, scalability, cost savings, high availability, and high fault tolerance. However, while cloud solutions offer tremendous opportunities and services to the industry, the landscape of cloud computing research is changing for several reasons, such as the emergence of data-intensive applications, multi-cloud deployment models, and more stringent non-functional requirements for cloud services. The proliferation of cloud computing providers on a global scale has surged significantly. Consequently, selecting an optimal provider that aligns with customer needs and defining appropriate evaluation criteria have become complex endeavors. Current trends reveal an increasing body of research focusing on appraising and ranking cloud-based services based on their ability to cater to user requirements. This paper synthesizes past investigations and recent studies, navigating the intricate landscape of modern cloud architectures. It delves into the dynamics of multiple providers, leveraging the advantages of decentralization. Additionally, it sheds light on the complexities of managing data storage and access within the dynamic context of multi-cloud environments, highlighting ongoing challenges. This comprehensive exploration culminates in a comparative summary of the existing approaches and a proposal for the forthcoming research, that includes a comprehensive set of criteria for the for multi-cloud data storage that involves a wide range of factors that can impact data placement, management, and retrieval across multiple cloud providers, contributing to the ongoing advancement of the field.

1 INTRODUCTION

As per the National Institute of Standards and Technology (NIST) interpretation, cloud computing can be defined as "a model of providing users with access to a shared pool of configurable computing resources, including networks, servers, storage, applications and services that can be rapidly provisioned (adapted for new services) with minimal management or interaction with service providers [1, 2, 3, 4]. In the same way, several sources state that Cloud Computing should be regarded solely as a way of remote access and management of computing resources [5]. The whole range of technologies that include Cloud Computing - these are already existing technologies, but as a way of combining them is a novelty [6]. To concentrate on the basic elements and functions of business, without worrying about maintenance or computer infrastructure and hiring specialized personnel, companies are increasingly introducing this technology into their business processes, which is essentially becoming another public service like electricity or the Internet [7]. It is necessary to highlight the main advantage of cloud computing, which is that due to a much more profitable use of computing resources in the cloud and their easy scalability, the results can also be obtained much faster. A good example of such an approach is comparing the cost of using one virtual machine for a long period of time (1000 hours) with the cost of using a large number of virtual machines (1000 VMs) for one hour. As a result, concurrently running machines can complete the task much faster, even though the monetary cost of these options may be identical.

In the following paper, we conduct the literature overview of existing approaches to managing distributed storage and access to Cloud data, review

their pros and cons, and establish a basis that allows developing a set of criteria and recommendations to distribute data between the multiple cloud providers to achieve the best result according to the defined criteria. The goals of re-examining the existing literature encompass three main aspects: firstly, to grasp the historical role of evaluation methods in quantifying the performance of cloud services. Secondly, to demonstrate the utility of applying specific approaches in the identification of suitable cloud vendors capable of meeting user requirements. Lastly, to accentuate unresolved difficulties and engage in discourse regarding future research directions.

In Section 2 we provide an overview of the main features of cloud computing and existing modern cloud architectures related particularly to storage and access. Next, Section 3 is analysing the challenges of heterogeneous multi-cloud systems from several perspectives. Section 4 provides an overview of existing research papers, and we discuss the highlights of each paper that we use for the comparison later. Section 5 discusses the lessons learned from the literature and identifies challenges and future research directions. We also define cloud evaluation criteria conducted from the previous research.

2 MODERN CLOUD ARCHITECTURES

Now being a prominent field of research in computer science, many scholars have proposed definitions and explanations of different aspects of cloud computing. Let us consider the main features of cloud computing [8]:

- It is available to users remotely at any time without the need for any additional direct hardware maintenance.
- Many available interfaces for use and access including cell phones, laptops, and desktops.
- Cloud providers focus on multi-tenant service delivery models, where the same physical resources are used together to provide services to multiple users at the same time.
- Flexibility and elasticity of service usage following end users' requirements.
- Availability of transparent and easily accessible service accounting.

Absence of Capex (Capital Expenditure) or simply upfront hardware investments and thus lower operating costs, the possibility of easy access and reduced network and hardware support tasks, allows small and medium-sized companies to focus on the core elements of the business and makes the choice of cloud computing a priority strategy for the near future [9].

Three models of cloud computing service delivery dominate the industry today: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS). In SaaS, customers can use the software on the provider's cloud system, usually over the Web [1, 4]. In PaaS, the provider allows the consumer to use the cloud network as a platform for their own developed or acquired software. In IaaS, virtual storage and machines are provided to the customer for greater control over the environment and the deployment of their applications [5].

Beyond basic considerations and specific use cases, the evolving nature of cloud solutions speaks to adoption at scale. Let us consider data storage: as data volumes increase exponentially, file-based storage now faces a challenge from new solutions, such as object-based frameworks, which empower the transfer of massive data sets at speed. Storage architecture is also evolving as options, such as hyper-converged and disaggregated, composable solutions emerge [10]. While hyper-converged offerings leverage software-defined clustering to provision storage resources on demand, disaggregated, composable systems decouple storage, network, and compute processes from physical hardware to create a shared pool of resources that can be individually assigned or used together. Open-source initiatives, meanwhile, are changing the fundamental nature of the cloud. From open-source software developments that allow organizations to customize key functions and frameworks to open-source architecture efforts-such as storage devices that are programmable and modifiable – the shift away from proprietary provisioning will impact long-term cloud decision-making. Organizations also must account for the movement of data and compute tasks away from central cloud services to edge infrastructure. Informed by increasingly complex processes handled by connected, intelligent devices at the point of origin-rather than being shunted to public or private clouds for analysis-edge computing modifies the value proposition of both public and private cloud services, as well as their impact on operational outcomes.

To better understand the motivation for multi-cloud, we segment the technical platform architecture into common scenarios for data distribution, by taking a suggested approach from [11] and further elaborating on it in terms of data management. The

key observations of the multi-cloud models could be described within the following categories. On the figures, each type of shape - squares, triangles, diamonds, and circles - serves as a symbol or visual identifier for the arbitrary data or objects for our case (can also represent applications or workloads).

Squares: within the Figures, squares are used as visual representations of arbitrary data objects or files. The use of squares might be arbitrary, indicating that the choice of this shape is not inherently tied to a specific meaning. In this context, squares could represent distinct parts of data objects or files, possibly signifying that different elements within a dataset share certain commonalities.

Triangles, Diamonds, and Circles: similar to squares, they serve as visual symbols for elements within the Figures. These triangles may represent various aspects of data, applications, or workloads. Like squares, they could be used to depict parts of a whole, where similar elements are represented using the same shape. This repetition suggests that these elements have shared attributes or characteristics.

The central idea here is that the choice of shape is flexible and not inherently tied to specific meanings. Instead, the emphasis is on visually distinguishing different elements within the data objects (applications or workloads). The use of the same shape, whether square, triangle, diamond, or circle, suggests that these elements have similarities or are parts of a larger entity, such as a data object or file. The figures aim to convey complex information or relationships through the visual representation of these shapes, allowing for clearer visualization and understanding of the data.

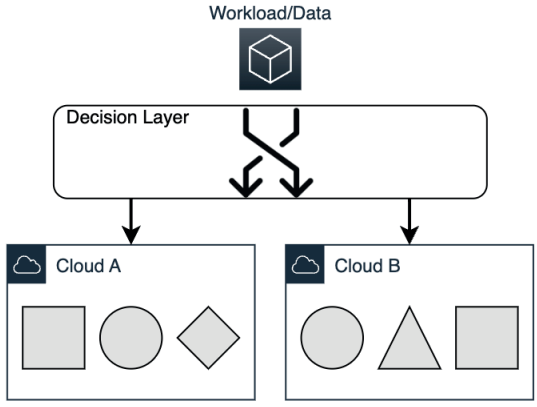


Figure 1: Random Data distribution scenario.

Figure 1 depicts a random scenario of a process for distributing data to a cloud provider. The decision layer, in this context, can be defined as an abstraction logic responsible for the selection of one cloud provider over another for data storage or deploying applications. The absence of robust governance and the influence of vendors are the primary factors leading to arbitrary decisions in choosing a cloud environment. Consequently, data is dispersed across multiple clouds, with some being processed in the first cloud, others in the second cloud, and additional portions in the third or fourth cloud. The lack of a well-defined process or criteria for determining where and what data should be stored contributes to this dispersed and less-than-optimal data distribution approach.

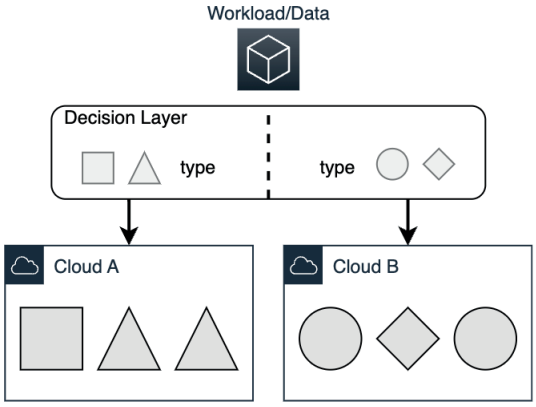


Figure 2: Segmented Data distribution scenario.

Figure 2 depicts segmenting data scenarios across different clouds, which is also common: specific types of data are being deployed to specific clouds. The following are the possible factors to decide on segmentation:

- Data Lifecycle (legacy or modern).
- Type of data (confidential vs. public).
- Type of product (compute vs. data analytics vs. collaboration software).

While considering this approach, we need to understand the traffic streams between clouds due to the possible excessive egress charges in case one half of the data ends up in one cloud and the other half in another one.

Often, the first two approaches might not be considered as true multi-cloud. The usual goal may be the ability to deploy data freely across cloud providers, thus minimizing vendor lock-in, usually by means of adding abstraction layers [12].

Figure 3 describes the choice scenario, which is common for large organizations' shared IT providers because they are expected to support a wide range of business units and their respective IT preferences. Often, such a setup involves a central commercial relationship and a common framework to create

instances on the cloud provider of choice but with corporate governance and constraints tacked on.

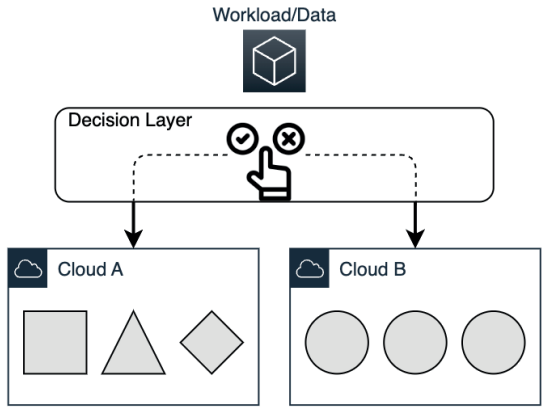


Figure 3: Choice Data distribution scenario.

The advantage of this setup is that projects are free to use proprietary cloud services, such as managed databases (depending on their preferred trade-off between avoiding lock-in and minimizing operational overhead). Hence, this setup makes a good initial step for "true" multi-cloud.

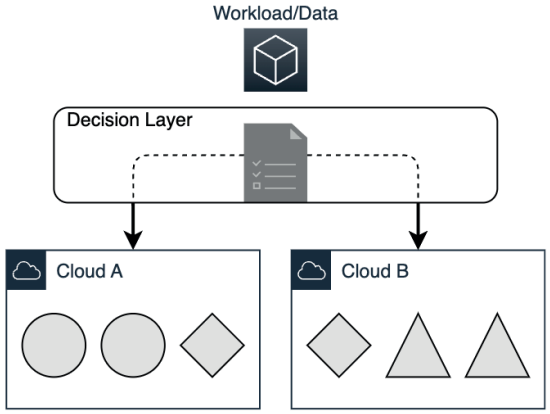


Figure 4: Parallel Data distribution scenario.

In Figure 4 we see the parallel data distribution approach, which corresponds to the notion of High Availability in the cloud, which is one of the main ensuring points for customers looking to store critical data across multiple clouds.

Being able to deploy the same application into multiple clouds requires a certain set of decoupling from the cloud provider's proprietary features. To achieve this, the following has to be considered [13]:

- Managing cloud-specific functions like identity management, deployment automation, or monitoring separately from the application in a cloud-specific manner.
- Using open-source components as much as possible – they will generally run on any cloud. However, it may reduce the ability to take advantage of other fully managed services, such as data stores or monitoring.
- Utilize a multi-cloud abstraction framework to perform a one-time development and be able to deploy to any cloud.
- Maintain two branches for those components of an application that are cloud provider-specific and wrap them behind a common interface. For example, there can be a common interface for block data storage.

The key aspect to watch out for is complexity, which can easily undo the anticipated uptime gain. Additional layers of abstraction and more tooling also increase the chance of a misconfiguration.

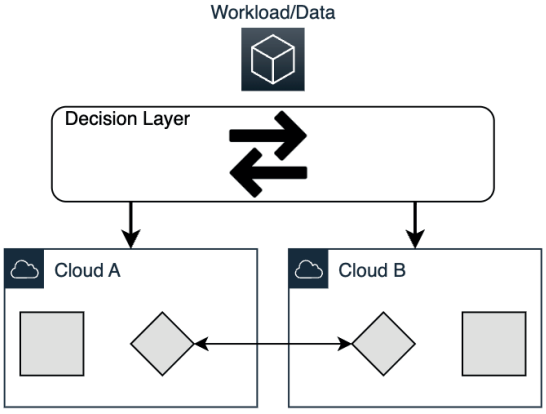


Figure 5: Portable Data distribution scenario.

Figure 5 depicts the best-case scenario for multi-cloud: free portability across cloud vendors. To the main benefits we assign avoiding vendor lock-in, and applications' placement based on resource needs. For example, day-to-day predictive and stable operations may be run in one cloud and burst excessive traffic into another.

The mechanism to enable this capability requires high levels of automation and abstraction away from cloud services. While for parallel deployments one could get away with a semi-manual setup or deployment process, full portability requires being able to shift the data at any time, so everything better be fully automated.

However, the cost of such a mechanism comes in the form of lock-in a specific vendor, product, and architecture plus a possible requirement for container distribution and orchestration [14]. Furthermore, these abstractions typically do not address data management, including the challenge of maintaining

data synchronization. Even if data synchronization is effectively handled, concerns about egress data costs may arise.

The following Table 1 summarizes the choices, the main drivers, and the side-effects to be aware of when choosing a particular scenario.

Table 1: Summary of the Multi-cloud distribution scenarios.

Common Challenges	Key Points to address in the research
Vendor Lock-In: In all scenarios, avoiding vendor lock-in is a primary concern. Organizations seek the flexibility to place their data across different cloud providers to minimize dependency on a single vendor.	Abstraction Logic: The need for a well-defined decision layer or abstraction logic is evident in all scenarios. This layer should facilitate the selection of cloud providers for data storage based on a comprehensive set of decision criteria. Research and development efforts should focus on creating a robust and vendor-agnostic decision layer.
Data Dispersal: A common challenge across the scenarios is the arbitrary distribution of data. Without well-defined processes and criteria, data ends up being dispersed across multiple clouds, leading to inefficiencies in data management and increased complexity.	Data Segmentation: Research should address the challenges of data segmentation, considering factors such as data lifecycle, data type, and the type of product. Efficient segmentation can help optimize data placement and reduce egress costs.
Egress Data Costs: The scenarios emphasize the potential for excessive egress charges when data is distributed across multiple clouds. Managing and optimizing data egress costs is a significant challenge, especially when data is scattered across different cloud environments.	High Availability: For scenarios involving parallel data distribution, research should explore methods for achieving high availability across multiple clouds. This includes managing cloud-specific functions, utilizing open-source components, and implementing multi-cloud abstraction frameworks while minimizing complexity.
Complexity: As organizations aim for multi-cloud solutions, they often encounter increased complexity. The need for additional layers of abstraction and tooling can lead to misconfigurations and operational challenges, potentially undermining the anticipated benefits of multi-cloud setups.	Free Portability: The best-case scenario involves free portability across cloud vendors. Research and development should focus on creating mechanisms for high levels of automation and abstraction, enabling seamless data migration and application deployment while addressing potential vendor lock-in and architectural considerations.

Next, we continue elaborating on the challenges of multi-cloud computing technology.

3 CHALLENGES OF THE MULTI-CLOUD SYSTEMS

Although cloud computing provides more flexible usage models for businesses by eliminating the need for dedicated allocation and allowing small businesses to scale their computing needs to meet business requirements, there are some issues and challenges facing cloud systems that may be the subject of future research [15]. Even though cloud computing is still in its continuous stage of development, its widespread adoption has allowed many researchers and companies to begin to assess the real and further potential challenges facing this technology [16].

Data is no longer just created in local data centers. The amount of data being created in the cloud and from emerging technologies - such as IoT and edge computing - continues to grow. Yet, according to various reports, companies are only capturing around 55% of the data potentially available through operations.

Capturing all available data, however, would overburden existing IT infrastructure and drive up costs. This is one of the many reasons why companies need to rethink their data management. If data is identified and classified at the beginning of its lifecycle, for example, this enables faster data cleansing, which in turn leads to lower costs.

Data proliferation results in silos that complicate the work of data scientists and analysts who transform this data into insights for decision-makers. Corporate culture can lead to further silos. Competing groups pursue their own goals and therefore want the ability to control and keep certain data to themselves.

To make data accessible from silos, business owners must overcome both technological and human barriers. Automated tools, such as Unified Policy Mechanisms, can solve the technological aspect. Global data management and global standards can help bring teams in line.

Data security is always a major concern for both IT and business management. Multi-cloud security comes with its own set of problems, such as inconsistent visibility across different clouds and a lack of coordination between different security components [17].

Vulnerable environments carry the risk of data breaches. The consequences of this range from financial losses and fines to reputational damage and data breaches. But the importance of security goes beyond this. Strong security is important to unlock the full value of data, ensuring unhindered access to data and data integrity.

Successful data management requires a holistic view of data storage, both in local and cloud architectures. This does not simply mean data democratization, but storage unification and data management through a centralized view, no matter where the data is stored.

The widespread simultaneous use of different storage technologies means a high space requirement, which can become a problem for companies. Moreover, there is often a lack of a coherent data storage strategy. Figure 6 depicts a typical approach to addressing the data definition question from 3 perspectives.

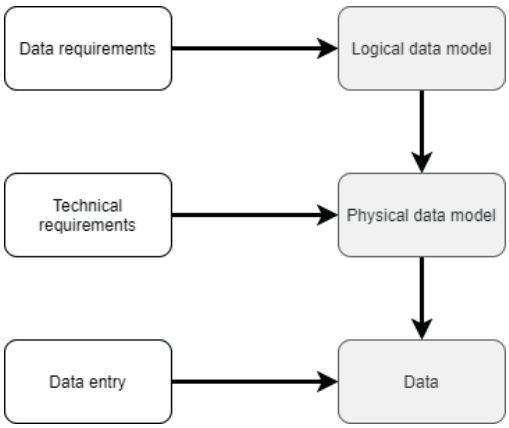


Figure 6: Holistic approach to data definition.

Often the data collected is stored in large repositories where it gathers dust. It is forgotten and businesses are missing out on a real treasure trove of information as a result. Smart data collection starts with understanding the business goals and insights that companies want to gain from their data. These goals make it clear what kind of data should be captured.

Organizing and sifting through massive amounts of data is also part of the data usability problem [18]. Companies must address factors such as complexity, overlapping tools, data integration, and more that impact the ability to obtain information valuable to the business.

A further challenge of data management in the cloud is the issue of steps and strategies for addressing possible catastrophic data loss. If, for example, a cloud provider goes insolvent or loses a data center due to a natural disaster, data may become unrecoverable. This would create a huge issue for large organizations that depend on data storage [18]. Compatibility or "the ability of different systems and organizations to work together (interoperate)" is the second most frequently mentioned cloud computing concern, after data security and trust issues. This is critical because it allows users to avoid vendor lock-in, in which the user becomes dependent on the Cloud Solution Provider (CSP) because they cannot migrate their services to other cloud platforms or clouds. The lack of standardization means that it can be difficult not only to transfer data between cloud systems but also to use them as one layer of data storage [19].

Another often-mentioned challenge is billing: pricing models are significantly different across providers, making it close to impossible to dynamically estimate the cost of the same task running on one cloud or another. To make the most of multi-cloud movement, enterprises need a three-step strategy:

1) Identify data architectures: before making the move to multi-cloud, it is critical for companies to identify current data architectures – including where data is stored, how it is used, and where processes could be improved – to define desired operational outcomes. With specific goals in mind, enterprises are better prepared to make multi-cloud decisions that deliver line-of-business benefits.
2) Assemble cloud resources: purpose-built clouds are now commonplace across both public and private deployments thanks to emerging solutions, such as customizable, composable infrastructure. As a result, it's critical for companies to assemble multi-cloud network, compute, memory, and storage functions that are capable of addressing current needs and meeting future challenges.
3) Enable seamless orchestration: multi-cloud deployments only deliver on performance and process mandates when cloud elements operate seamlessly. Realizing Return of Investment (ROI) on this move to multi-cloud means building on defined architectures and assembled resources with robust data orchestration and management solutions capable of reducing friction across key cloud functions.

One of the pivotal challenges that cloud computing users face is selecting the most efficient and cost-effective data storage and access methods while circumventing vendor lock-in. Nonetheless, the extant criteria for optimizing data distribution in multi-cloud environments only encompass a subset of customer requirements. Despite prior efforts to delineate these criteria, the rapidly evolving cloud computing landscape necessitates a comprehensive reevaluation of approaches [20].

4 LITERATURE OVERVIEW

This section investigates the different solutions related to data storage and retrieval in a multi-cloud environment.

Yang and Ren [21] proposed a Virtual Framework for Cloud Storage Services (VCSS) to integrate diverse common open Cloud Storage Services to be a uniform virtual storage resources pool. Key concepts and technologies include a service metadata model with rich NFP, a service repository serving as a virtual storage resource pool to enable resources physically distributed stored, and logically centralized managed, and a service schedule model serving to estimate and select an appropriate service to store and replicate the file.

Megouache et al. [22] presented a novel module for solving security problems in multi-cloud platforms. This method contains 3 phases, the initial stage, is for proposing a private virtual network to secure the data transmission. Next, they utilized an authentication technique depending upon data encryption, for protecting the user's identity and information, and lastly, they created a method for knowing the reliability of data allocated on the several clouds of the scheme. The module attains identity verification and the capability to interoperate among processes run on distinct cloud providers. A data integrity method would also be established.

Colombo [23] encrypted the file as a whole and gave a choice to the user to select the cloud provider among the hybrid cloud. Irrespective of the sensitivity of data stored in the cloud, the entire file is encrypted. Also, to provide fine-grained access to data, multiple copies of data are kept in the cloud. High retrieval overhead is a problem in this solution.

Subramanian and Leo [24] aim to provide a framework that decreases malicious insiders and file risks and enhances data sharing security in multi-cloud storage services. This method would provide a secured platform where the data owner could retrieve and store data from the multi-cloud platform with no merging file conflict and prevent insider attacks from obtaining useful data. Research indicates that the recommended module is appropriate for making decision procedures for the data owners in an optimum acceptance of multi-cloud storage service to share their data safely.

Cao [25] implemented and designed a multi-cloud architecture to build Open Stack based environment for medicinal IoT, denoted by Tri-SFRS. For implementing this technique, they integrate various methods for attaining this decrease in efforts, comprising lower overhead native testing architecture, multi-cloud cascading framework, snapshot volume cascaded operation for b-ultrasonic data, and medicinal data storage backup method. Tri-SFRS can concurrently allow asset management. Tri-SFRS was implemented as a native element in the Open Stack environment, and it determines the degree of native Open Stack multi-cloud environment management using this presented cascading architecture.

Celesti [26] deliberated to improve the whole system regarding retrieval and data storage via validating and testing an MCS scheme consisting of 3 main Cloud Storage suppliers: Copy, Dropbox, and Google Drive. Research has shown that the selection of a Cloud storage provider for storing files according to data transfer efficiency depends on file chunk size.

Samundiswary [27] proposed an object storage architecture for unstructured data. Metadata about the object is used for searching in the object storage architecture. Though the object storage architecture proposed in this work assists in rapid retrieval, it is not secure against attacks.

Libardi [28] proposed a Multi-cloud Storage Selection Framework to automatically select a storage dispersal strategy. MSSF formalizes the selection process using a knapsack optimization problem using integer linear programming along with a rule-based system to select a multi-cloud storage strategy that fits the user's needs and requires only simple inputs from the user. Its main limitation is the requirement of user input parameter selection for each file upload.

Li et al. [29] proposed a privacy-preserving STorage and REtrieval (STRE) method that guarantees privacy and security however also offers consistency assurances for the outsourced searchable encrypted data. The STRE method allows the cloud user to distribute and search its encryption data over many independent clouds handled by distinct CSPs, and strong while a specific amount of CSP crashes. In addition to reliability, STRE provides the advantage of partly hidden search pattern. Though this scheme is secure and privacy-preserving, overhead of Shamir splitting and reassemble is very high for large data volume. With only a small percentage of sensitive information in the unstructured data, this overhead is too high.

Janviriya [30] approach consists of Multiple Cloud Storage Integration Systems based on RAID (Redundant array of independent disks) 0 stripping technology where the proposed application creates a single cloud storage combined out of multiple cloud storage accounts and also decreases the total time of file uploading and downloading to and from the

cloud. Also, because of concerns about security on the cloud, the research suggests the security level enhancement of its cloud-stored files. The results from experiments using the application show an improvement in performance, storage capacity, and security.

Zhao et al. [31] proposed a middleware that enables any end-user application to automatically and securely store files in multiple cloud storage accounts. middleware that enables an application to use the cloud storage services securely and efficiently, without any code modification or recompilation. The proposed solution securely saves the user data to different cloud storage services to significantly enhance the data security, without the need to save it to the local disk. The solution is implemented as a shared library on Linux and supports applications written in different languages, supports various popular cloud storage services, and supports common user authentication methods used by those services.

Rios [32] proposed a new DevOps architecture intended to assist Cloud consumers in deploying, designing, and functioning (multi) Cloud systems that contain the required security and privacy controls to ensure law enforcement authorities, transparency for end users, and third-party in-service provisions. The architecture is based on the risk-driven requirement at the implementation time of security and privacy levels objective in the continuous enforcement and service level agreement and observing at run-time.

Tchernykh et al. [33] presented a multi-cloud-based storage framework named WA-RRNS which integrates threshold secret allocation redundant and weight access system remains number scheme with many failure recognition or recovery mechanism and homomorphic cipher. For optimum trade-offs between security and efficiency, WA-RRNS utilizes variables for adjusting data loss probability, redundancy and encryption/decryption speed. Investigational and Theoretical analyses with actual data displays that this method gives a secure manner for mitigating the uncertainty of untrusted and not consistent cloud storage.

Pravin [34] addressed the privacy and security risks of data in the multi-cloud storage. The proposed solution is based on the cryptographic technique with a dynamic file-slicing method for securing data in a multi-cloud environment. The stored data is fragmented into many slices. The number of slices is defined by the data owner. The sliced files are encrypted with 3 DES (data encryption standard) and elliptical curve cryptography (ECC) algorithm. The performance of the proposed technique was evaluated using latency time and the results revealed that the proposed technique outperformed the other methods discussed in the article.

Esposito [35] proposed three different methods for selecting the best cloud service set in order to maximize the quality of service and minimize cost. The three methods utilized are based on Fuzzy Logic, Theory of Evidence and Game Theory. The suggested solution mainly focuses on availability and cost.

Le [36] proposed a data partitioning model based on fragmentation, secret sharing, and encryption for medical data storage in clouds. Patient-centric information is represented as an Entity association and relation model. Access control to the sensitive information in the patient-centric model is enforced using cryptographic algorithms.

Valapula [37] proposed a secure and efficient data partitioning scheme for a hybrid cloud is proposed. The scheme is adaptive to unstructured text documents and achieves a fine balance between multiple objectives of maximizing security, increasing public cloud utilization, and minimizing the degradation of retrieval efficiency. The performance of the solution was tested against the Enron email dataset.

Vernik [38] presented an on-boarding federation mechanism for adding a special layer on cloud storage services allowing them to import data from other services. This was achieved without dependency on special functions from the other cloud vendors. The proposed solution suggested a design of a generic, modular onboarding architecture designed for content-centric data. This approach requires certain adaptability level from the cloud service provider becoming hard to implement. However, no fault tolerance mechanism is proposed.

Malensek [39] proposed a private and public cloud federation method in order to improve queries throughput in large datasets. Their distributed storage framework autonomously tuned in-memory data structures and query parameters to ensure efficient retrievals and minimize resource consumption. To avoid processing hotspots, they predicted changes in incoming traffic and federated their query resolution structures to the public cloud for processing. The efficacy of the suggested frameworks was demonstrated on a real-world, petabyte dataset. In addition, several approaches referred to Service Selection and Data Distribution were studied.

Sukmana [40] proposed a unified cloud access control model that provides the abstraction of Cloud Service Providers (CSP) services for centralized and automated cloud resource and access control management in multiple CSPs. Their proposal offered

role-based access control for Cloud Storage Brocker (CSB) stakeholders to access cloud resources by assigning necessary privileges and access control lists for cloud resources and CSB stakeholders, respectively, following the privilege separation concept and least privilege principle.

Chang [41] suggested a mathematical formulation of the cloud service provider selection problem in which both the object functions and cost measurements are defined. The algorithms that are selected among cloud storage providers to maximize the data survival probability or the amount of surviving data are subject to a fixed budget, and a series of experiments demonstrated that the proposed algorithms were efficient enough to find optimal solutions in a reasonable amount of time, using price and fail probability taken from real cloud providers.

5 CRITERIA FOR MILTI-CLOUD

Based on the conducted literature review and modern cloud computing standards for storage and access we define a complex set of criteria for multi-cloud data storage that involves considering a wide range of factors that can impact data placement, management, and retrieval across multiple cloud providers. Table 2 depicts a comprehensive set of criteria to consider in the future research.

Future work should incorporate multi-cloud paradigms and involve an intricate exploration of each criterion, further refining their definitions and relevance within the ever-evolving cloud computing landscape. Initially, based on the suggested set of criteria we define a comprehensive framework for multi-cloud data storage strategies, encompassing technical, operational, and business considerations. Subsequently, we establish an ontology, serving as an abstract logical layer for the distribution of data based on the predefined set of criteria.

Following this, we devise an algorithm for the optimal allocation and storage of data, leveraging the ontology model created in the previous step. This algorithm ensures consideration of multiple criteria in alignment with the established ontology.

The final step involves developing a method for organizing optimal data access, taking into account the distributed and stored data following the predefined criteria. Concurrently, we introduce a pseudo-query language tailored for data retrieval, capable of describing algorithms for data retrieval based on selected operations. These operations may encompass full or partial data retrieval and filtering by date or data type, among other criteria.

Table 2: Complex set of criteria for multi-cloud data storage distribution.

#	Criteria Category	Specific Criteria	Possible Measurement Metric
1	Data Accessibility Criteria	Latency Requirements	Milliseconds (ms)
2		Redundancy and Availability	Availability Percentage (%)
3		Data Consistency	Data Consistency Index
4		Data Encryption	Encryption Strength (e.g., AES-256)
5	Cost and Resource Utilization Criteria	Cost Efficiency	Cost per GB/month ($)
6		Resource Allocation	Resource Utilization (%)
7		Data Lifecycle Management	Percentage of Archived Data (%)
8	Data Type and Format Criteria	Data Classification	Data Classification Score
9		Data Format	Data Format Compatibility
10	Compliance and Security Criteria	Regulatory Compliance	Compliance Audit Score
11		Data Ownership	Data Ownership Policy Adherence
12		Security Protocols	Security Protocol Strength
13	Scalability and Performance	Scalability	Scalability Factor
14		Performance Metrics	Throughput (requests/second)
15	Data Migration and Interoperability Criteria	Data Portability	Data Portability Index
16		Interoperability	Interoperability Score
17	Vendor Lock-In and Vendor Criteria	Vendor Lock-In Mitigation	Lock-In Reduction Score
18		Vendor Reputation	Vendor Reputation Rating
19	Disaster Recovery and Backup	Recovery Time Objective (RTO)	Recovery Time Objective (RTO, hours)
20		Recovery Point Objective (RPO)	Recovery Point Objective (RPO, hours)
21		Data Backup Frequency	Frequency (e.g., per day, per week)
22		Backup Storage Redundancy	Redundancy Level (e.g., dual-site)
23	Monitoring and Reporting	Monitoring Tools	Tool Effectiveness (e.g., Score)
24		Reporting	Reporting Accuracy (e.g., Percentage)
25	Sustainability	Environmental Impact	Carbon Emission Reduction (%)
26		Energy Efficiency	Energy Usage (kWh)
27		Resource Sustainability	Resource Conservation Index

These research goals are strategically designed to systematically address the challenges of information environment representation, efficient data allocation and storage, and streamlined data access in a structured and methodical manner.

6 CONCLUSIONS

In summary, cloud computing remains a rapidly evolving technology with diverse applications across various industries, particularly in remote computing and storage. However, it is important to acknowledge that there are still many unresolved issues and untapped possibilities in this field.

In this paper, we have revisited the existing challenges associated with multi-cloud solutions and conducted a comprehensive review of solutions proposed in scientific literature. To address these challenges and expand the scope of potential solutions, we have introduced a comprehensive set of criteria that will serve as the foundation for future research. Additionally, we have outlined the primary directions for our upcoming research and highlighted the expected outcomes.

Our overarching goal is to address the complex challenges of multi-cloud data distribution, ultimately enhancing performance, minimizing costs, and ensuring data accessibility while catering to the diverse demands of customers.

REFERENCES

[1] J. Hong, T. Dreibholz, J. Schenkel, and J. Hu, "An Overview of Multi-cloud Computing," In: Proceedings of the International Conference on Cloud Computing, 2019. [Online]. Available: https://doi.org/10.1007/978-3-030-15035-8_103.

[2] "Products in Cloud Infrastructure and Platform Services." Gartner, [Online]. Available: https://www.gartner.com.

[3] J. Alonso, L. Orue-Echevarria, V. Casola, et al., "Understanding the Challenges and Novel Architectural Models of Multi-cloud Native Applications – A Systematic Literature Review," Journal of Cloud Computing, vol. 12, no. 6, 2023. [Online]. Available: https://doi.org/10.1186/s13677-022-00367-6.

[4] M. Peter and G. Tim, "The NIST Definition of Cloud Computing," National Institute of Standards and Technology, Tech. Rep., 2011.

[5] T.B. Winans and J.S. Brown, "Cloud Computing: A Collection of Working Papers," Deloitte LLC, 2009.

[6] Z. Qi, "Cloud Computing: State-of-the-Art and Research Challenges," Journal of Internet Services and Applications, vol. 1, no. 1, 2010.

[7] F. Armando, G. Rean, J. Anthony, K. Randy, K. Andrew, et al., "Above the Clouds: A Berkeley View of Cloud Computing," Tech. Rep. UCB/EECS2009-28, EECS Department, University of California, Berkeley, 2009.

[8] R. Buyya, S.N. Srirama, G. Casale, R. Calheiros, Y. Simmhan, et al., "A Manifesto for Future Generation Cloud Computing: Research Directions for the Next Decade," ACM Computing Surveys, vol. 51, no. 5, 2018.

[9] N. Antonopoulos and L. Gillam, "Cloud Computing: Principles, Systems, and Applications," London: Springer, 2010.

[10] D. Slamanig and C. Hanser, "On Cloud Storage and the Cloud of Clouds Approach," In: Proceedings of the International Conference on Internet Technology and Secured Transactions, 2012. [Online]. Available: https://doi.org/10.1109/ICITST.2012.6470979.

[11] G. Hohpe, "Multi Cloud Architecture: Decisions and Options," July 2019.

[12] Y. Ghanam, J. Ferreira, and F. Maurer, "Emerging Issues and Challenges in Cloud Computing – A Hybrid Approach," Journal of Software Engineering and Applications, vol. 5, no. 11A, 2012.

[13] D. Petcu, "Portability and Interoperability between Clouds: Challenges and Case Study," In: Towards a Service-Based Internet. Springer, Berlin/Heidelberg, 2011.

[14] T. Dreibholz, "Big Data Applications on Multi-Clouds: An Introduction to the MELODIC Project," Keynote Talk at Hainan University, College of Information Science and Technology, 2017.

[15] Y. Elkhatib, "Mapping Cross-Cloud Systems: Challenges and Opportunities," In: Proceedings of the 8th USENIX Conference on Hot Topics in Cloud Computing. Berkeley/United States, 2016.

[16] O. Tomarchio, D. Calcaterra, and G.D. Modica, "Cloud Resource Orchestration in the Multi-Cloud Landscape: A Systematic Review of Existing Frameworks," Journal of Cloud Computing, vol. 9, no. 49, 2020. [Online]. Available: https://doi.org/10.1186/s13677-020-00194-7.

[17] N. Vurukonda and B.T. Rao, "A Study on Data Storage Security Issues in Cloud Computing," Procedia Computer Science, vol. 92, 2016.

[18] T.G. Papaioannou, N. Bonvin, and K. Aberer, "Scalia: An Adaptive Scheme for Efficient Multi-Cloud Storage," In: Proceedings of the International Conference on High-Performance Computing, Networking, Storage, and Analysis, 2012. [Online]. Available: https://doi.org/10.1109/SC.2012.20.

[19] D. Petcu, "Consuming Resources and Services from Multiple Clouds," Journal of Grid Computing, vol. 12, pp. 321–345, 2014. [Online]. Available: https://doi.org/10.1007/s10723-013-9290-3.

[20] S. Bharany, S. Sharma, O.I. Khalaf, et al., "A Systematic Survey on Energy-Efficient Techniques in Sustainable Cloud Computing," Sustainability, vol. 14, no. 10, 2022. [Online]. Available: https://doi.org/10.3390/su14105712.

[21] D. Yang and C. Ren, "VCSS: An Integration Framework for Open Cloud Storage Services," In: Proceedings of the IEEE World Congress on Services, 2014. [Online]. Available: https://doi.org/10.1109/SERVICES.2014.36.

[22] L. Megouache, A. Zitouni, and M. Djoudi, "Ensuring User Authentication and Data Integrity in Multi-Cloud Environment," Human-centric Computing and Information Sciences, vol. 10, 2020. [Online]. Available: https://doi.org/10.1186/s13673-020-00254-7.

[23] M. Colombo, R. Asal, Q.H. Hieu, et al., "Data Protection as a Service in the Multicloud Environment," In: Proceedings of the IEEE 12th International Conference on Cloud Computing (CLOUD), 2019.

[24] K. Subramanian and J. Leo, "Enhanced Security for Data Sharing in Multi-Cloud Storage (SDSMC)," International Journal of Advanced Computer Science and Applications, vol. 8, 2017.

[25] R. Cao, Z. Tang, C. Liu, and B. Veeravalli, "A Scalable Multicloud Storage Architecture for Cloud-Supported Medical Internet of Things," IEEE Internet of Things Journal, vol. 7, no. 3, 2019.

[26] A. Celesti, A. Galletta, M. Fazio, and M. Villari, "Towards Hybrid Multi-Cloud Storage Systems: Understanding How to Perform Data Transfer," Big Data Research, vol. 16, 2019.

[27] S. Samundiswary and N.M. Dongre, "Object Storage Architecture in Cloud for Unstructured Data," In: International Conference on Inventive Systems and Control (ICISC), 2017.

[28] R.M.d.O. Libardi, S. ReiMarganiec, L.H. Nunes, et al., "MSSF: User-Friendly Multi-Cloud Data Dispersal," In: Proceedings of the IEEE 8th International Conference on Cloud Computing, 2015.

[29] J. Li, D. Lin, A.C. Squicciarini, J. Li, and C. Jia, "Towards Privacy Preserving Storage and Retrieval in Multiple Clouds," IEEE Transactions on Cloud Computing, vol. 5, no. 3, 2017. [Online]. Available: https://doi.org/10.1109/TCC.2015.2485214.

[30] P. Janviriya, T. Ongarjithichai, P. Numruktrakul, and C. Ragkhitwetsagul, "CloudyDays: Cloud Storage Integration System," In: Proceedings of the Third ICT International Student Project Conference (ICT-ISPC), 2014. [Online]. Available: https://doi.org/10.1109/ICT-ISPC.2014.6923233.

[31] R. Zhao, C. Yue, B. Tak, and C. Tang, "SafeSky: A Secure Cloud Storage Middleware for End-User Applications," In: Proceedings of the IEEE 34th Symposium on Reliable Distributed Systems (SRDS), 2015. [Online]. Available: https://doi.org/10.1109/SRDS.2015.23.

[32] E. Rios, E. Iturbe, X. Larrucea, M. Rak, W. Mallouli, et al., "Service Level Agreement-Based GDPR Compliance and Security Assurance in (Multi) Cloud-Based Systems," IET Software, vol. 13, no. 3, 2019.

[33] A. Tchernykh, M. Babenko, V. Miranda-López, A.Y. Drozdov, and A. Avetisyan, "WA-RRNS: Reliable Data Storage System Based on Multi-Cloud," In: Proceedings of the IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), 2018. [Online]. Available: https://doi.org/10.1109/IPDPSW.2018.00124.

[34] A. Pravin, T.P. Jacob, and G. Nagarajan, "Robust Technique for Data Security in Multicloud Storage Using Dynamic Slicing with Hybrid Cryptographic Technique," Journal of Ambient Intelligence and Humanized Computing, 2019.

[35] C. Esposito, M. Ficco, F. Palmieri, and A. Castiglione, "Smart Cloud Storage Service Selection Based on Fuzzy Logic, Theory of Evidence and Game Theory," IEEE Transactions on Computers, vol. 65, no. 8, 2016. [Online]. Available: https://doi.org/10.1109/TC.2015.2389952.

[36] D.-N. Le, B. Seth, and S. Dalal, "A Hybrid Approach of Secret Sharing with Fragmentation and Encryption in Cloud Environment for Securing Outsourced Medical Database: A Revolutionary Approach," Journal of Cyber Security and Mobility, vol. 7, 2018.

[37] S. Vulapula and H. Valiveti, "Secure and Efficient Data Storage Scheme for Unstructured Data in Hybrid Cloud Environment," Soft Computing, vol. 26, 2022. [Online]. Available: https://doi.org/10.1007/s00500-022-06977-1.

[38] G. Vernik, A. Shulman-Peleg, S. Dippl, C. Formisano, M.C. Jaeger, et al., "Data On-boarding in Federated Storage Clouds," In: Proceedings of IEEE International Conference on Cloud Computing CLOUD, 2013.

[39] M. Malensek and S. Pallickara, "Autonomous Cloud Federation for High-Throughput Queries Over Voluminous Datasets," IEEE Cloud Computing, vol. 3, no. 3, 2016. [Online]. Available: https://doi.org/10.1109/MCC.2016.65.

[40] M.I.H. Sukmana, K.A. Torkura, H. Graupner, F. Cheng, and C. Meinel, "Unified Cloud Access Control Model for Cloud Storage Broker (PS23)," In: Proceedings of the International Conference on Information Networking (ICOIN), 2019.

[41] C.W. Chang, P. Liu, and J.J. Wu, "Probability-Based Cloud Storage Providers Selection Algorithms with Maximum Availability," In: Proceedings of the 41st International Conference on Parallel Processing, 2012. [Online]. Available: https://doi.org/10.1109/ICPP.2012.51.

Comparing Classical Machine Learning and Deep Learning for Classification of Arrhythmia from ECG Signals

Marija Bikova, Vesna Ojleska Latkoska and Hristijan Gjoreski
Faculty of Electrical Engineering and Information Technologies, "Ss. Cyril and Methodius University" in Skopje, Rugjer Boshkovikj 18, Skopje, North Macedonia
marija_bikova@yahoo.com, vojleska@feit.ukim.edu.mk, hristijang@feit.ukim.edu.mk

Keywords: Cardiac Arrhythmia, Deep Learning, Classification Electrocardiogram, Convolutional Neural Network, Long- Short Term Memory.

Abstract: Arrhythmia detection is a vital task for reducing the mortality rate of cardiovascular diseases. Electrocardiogram (ECG) is a simple and inexpensive tool that can provide valuable information about the heart's electrical activity and detect arrhythmias. However, manual analysis of ECG signals can be time-consuming and prone to errors. Therefore, machine learning models have been proposed to automate the process and improve the accuracy and efficiency of arrhythmia detection. In this paper, we compare six machine learning models, namely ADA boosting, Gradient Boost, Random Forest, C-Support Vector (SVC), Convolutional Neural Network (CNN), and Long Short-Term Memory Network (LSTM), for arrhythmia detection using ECG data from the MIT-BIH Arrhythmia Database. We evaluate the performance of the models using various metrics, such as accuracy, precision, recall, and F1-score, on different classes of ECG beats. We also use confusion matrices to visualize the errors made by the models. We find that the CNN model is the best performing model overall, achieving accuracy of 95% and F1-score of 84.75%. SVC and LSTM were the second and third best, achieving accuracy of 94% and 93%, respectively. We also discuss the challenges of using ECG data for arrhythmia detection, such as noise, imbalance, and similarity of classes. We suggest some possible ways to overcome these challenges, such as using more advanced preprocessing and resampling techniques, or incorporating domain knowledge and expert feedback into the models.

1 INTRODUCTION

Cardiovascular diseases (CVDs) have been the leading cause of death since 1999 as the statistics of the Centers for Disease Control and Prevention indicate [1]. The mortality rate can be effectively reduced by providing a timely treatment using a classification model to identify CVDs at early stage [7]. One of the common sources of CVDs is cardiac arrhythmia, where heartbeats are known to deviate from their regular beating pattern. A normal heartbeat varies with age, body size, activity, and emotions. In cases where the heartbeat feels too fast or slow, the condition is known as palpitations. An arrhythmia does not necessarily mean that the heart is beating too fast or slow, it indicates that the heart is following an irregular beating pattern. It could mean that the heart is beating too fast-tachycardia, when there are more than 100 beats per minute (bpm), or slow – bradycardia with less than 60 bpm, skipping a beat, or in extreme cases, cardiac arrest. Some other common types of abnormal heart rhythms include atrial fibrillation, atrial flutter, and ventricular fibrillation [2]. The Electrocardiogram (ECG) signal detects cardiac abnormalities by measuring the electrical signals generated by the heart during contraction. A careful study of ECG signals is crucial for precise diagnoses of patients' acute and chronic heart conditions. Arrhythmia is a cardiac abnormality related to the rate and rhythm of the heartbeat [6]. Despite being the most frequently used diagnosing tool, the rates of ECGs misdiagnosis are still too high. It is very challenging to accurately detect the clinical condition presented by an ECG signal. Cardiologists need to accurately predict and identify the right kind of abnormal heartbeat ECG wave and then recommend the appropriate treatment. The analysis of the electrocardiogram (ECG) signals is done manually which can be time-consuming. To address this issue, machine learning (ML) classification is being

proposed to automate the process. This would allow ML models to learn the features of a heartbeat and detect abnormalities [10][11].

2 ECG STRUCTURE AND MIT-BIH DATABASE

The human body can be thought of as a giant conductor of electrical currents. An electrocardiogram (ECG) can be registered by connecting electrical leads to any two points on the body. The ECG contains records for the electrical activity of the heart. The ECG of the heart forms a series of waves and complexes that have been labelled in alphabetical order: the P wave, the QRS complex, the T wave and the U wave. The P wave is produced by depolarization of the atria; depolarization of the ventricles produces the QRS complex; and repolarization of the ventricles causes the T wave [3]. The significance of the U wave is uncertain. Each of these electrical stimulations results in a mechanical muscle twitch. This is called the electrical excitation-mechanical contraction coupling of the heart. This allows us to detect abnormalities by equating each phase to the normal cardiac cycle. Figure 1 shows the ECG signal representation of a normal beat. These ECG signals are extremely susceptible to high and low frequency noise which usually occur from baseline wander, misplaced electrode contact, motion artifacts, or power line interference [3].

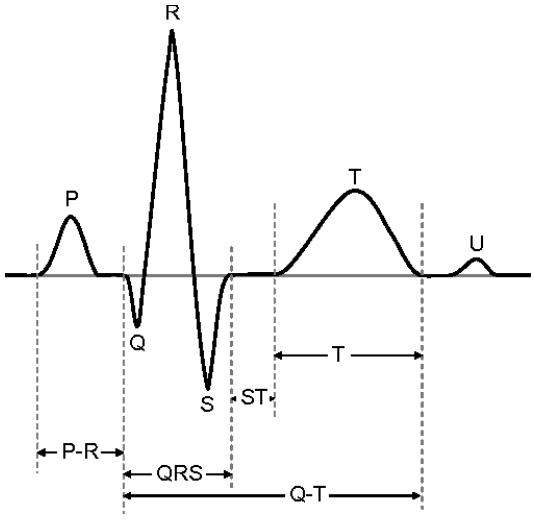


Figure 1: Electrocardiogram (ECG), showing significant intervals and deflections [3].

The MIT-BIH Arrhythmia Database [12] is a publicly available database that contains sections of ambulatory ECG recordings from 47 subjects. The recordings were digitized at 360 samples per second per channel with 11-bit resolution at 10-mV range on two channels and studied by the BIH Laboratory. Here, 23 recordings were picked at random from a set of 4000 24-hour ECG recordings collected from a population of 60% inpatients and 40% outpatients [12]. The dataset has been pre-annotated and labelled by cardiologists. These different annotations refer to various normal and abnormal ECG signals which represent different types of arrhythmia. The dataset consists of ECG signals of various classes, but the eight classes used for this investigation are 'N', 'L', 'R', 'V', 'A', 'F', 'f', and '/'.Table 1 shows the description and numerical identification values assigned to these classes [3].

Table 1: Beat classes, ID number and description.

Class	ID	Beat Description
N	1	Normal
L	2	Left Bundle Branch Block
R	3	Right Bundle Branch Block
V	4	Premature Ventricular Contraction
A	5	Atrial Premature
F	6	Fusion of Ventricular and Normal
f	7	Fusion of Paced and Norma;
/	8	Paced

3 DATA PREPARATION

We used the pre-processing methodology as proposed by Verma et al. [3]. The MIT-BIH dataset was read using the native python waveform-database (WFDB) package, a library of tool for reading, writing, and processing WFDB signals and annotations. Most of the ECG signals were assigned to the annotation classes explained in Table 1. The ECG dataset is imbalanced, since there is an abundance on 'N' beats and the other beat classes do not pass the 10000 thresholds. This is only from one channel of the MIT-BIH database. To get all the beats we extracted and stacked the ECG signals from both the channels. After extracting the 8 classes that are going to be used and removing the other classes, a data clean-up processes were applied. First, the data was made purely numerical in order for easier working, by assigning each of the 8 classes a number, shown in Table 1.

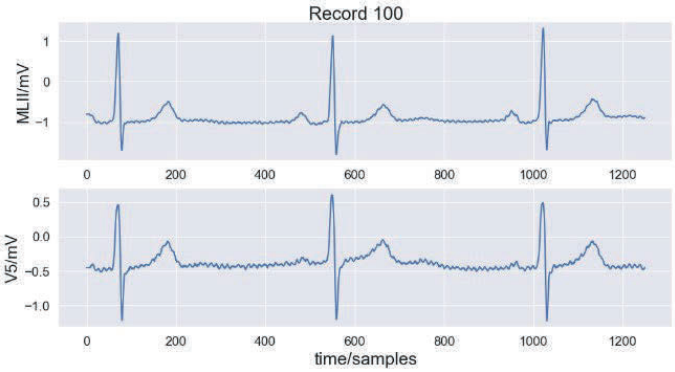


Figure 2: Example of original ECG beats from MIT-BIH database (X-axis: Timestamps, Y-axis: Voltage).

The next step was to make every beat contain equal amount of data points, so each individual beat was extracted from all the records by matching the R-peaks of the ECG with the respective annotation class and appending the class numerical value at the end of the beat. Then standardization process was implemented for all beats to ensure consistent signal amplitudes, using the formula $z=(x-\mu)/\sigma$, where the new beat is represented with z (x is the original beat data, μ is the mean of the beat data and σ is the standard deviation of the data). Each beat was labeled with the patient record number and the annotation class number, and the resulting clean data was then saved into a single .csv file, containing all beats from all records.

The dataset with the clean data was divided into two fundamental subsets, train and test. The test set represented 25% of the original dataset, while the remaining 75% formed the training set.

The original data from MIT-BIH database for one patient is shown on Figure 2. It can be noticed that the ECG signals are continuous and not standardized between the two channels and are sampled at 360 Hz. The data after performing pre-processing is shown in Figure 3, where the standardization and r-peak centering that is implemented on the data points can be seen.

The resample technique by Sci-kit Learn is used in to address the imbalance between the classes in the MIT-BIH dataset. The bootstrap method is involved in this technique of resampling, where statistics are estimated on a data population by sampling a dataset with replacement through iteration using a sample size and number of repeats. By taking the mean values of the total number of beats of the abnormal classes the value for up-sampling and down-sampling denoted as n_samples were calculated. After resampling, all eight classes in the training dataset have 3989 samples for the beat hold out method.

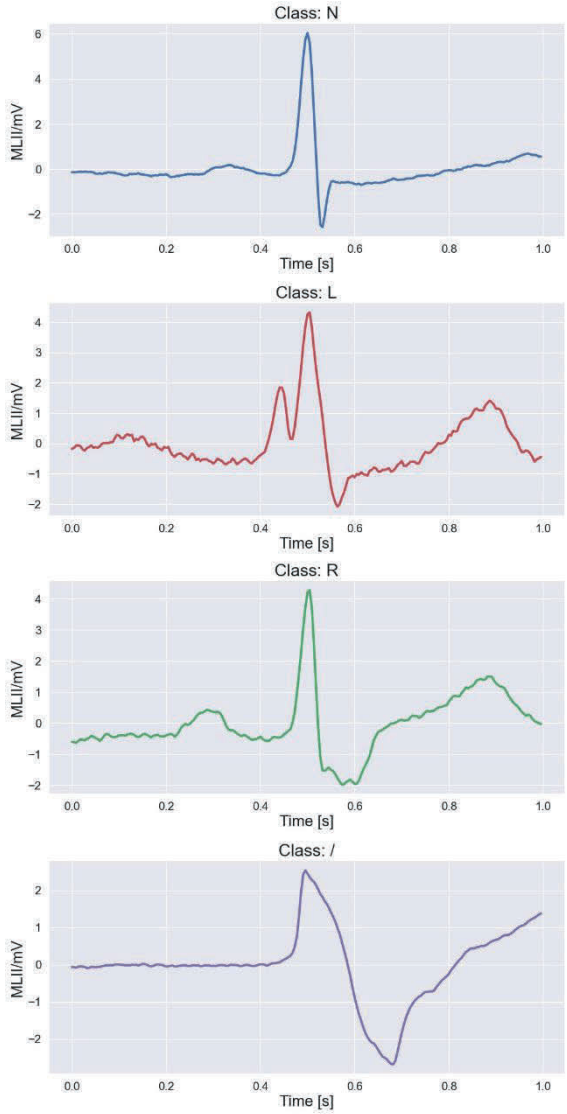


Figure 3: Example of single beats of N, L, R and / classes (X-axis: Timestamps, Y-axis: Voltage).

4 MACHINE LEARNING MODELS

We use 6 models that have different characteristics to classify ECG data into multiple categories based on their patterns. We decided to use deep learning models, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, because of their ability to capture spatial features and temporal dependencies, respectively. This choice was supported by many previous research successes in similar contexts. We also employed classical machine learning models, such as Gradient Boosting (GBC), ADA Boosting (ADA), Random Forecast (RFC), C-Support Vector (SVC), which are well-known for their performance in various classification tasks.

Prior successful research papers [3][4] guided us through the selection of our models. Verma et al. [3], proposed an 11-layer CNN model and LSTM models to classify 8 classes of beats in the MIT-BIH arrhythmia dataset and their models displayed an accuracy of 94.1% and 94% for K-Fold cross-validation method, and 98.7% and 97% for Leave Groups Out method, respectively. Their CNN model had four layers of 1D-convolutiona and batch normalization pairs, with ReLU activation, 16 kernel size and 128, 32 filters. The final two layers were 1D-convolution layer with 9 filters and a 1D-max pooling layer with 2 pool size, followed by a flatten layer and fed to four dense layers. The LSTM model consisted of two LSTM layers with 128 and 9 filters, followed by a 1D-max pooling layer with 2 pool size. The output is flattened and goes to four dense layers with ReLU and softmax activations.

Pandey and Janghel et al. [4] used an 11-layer CNN with SMOTE to classify five beat classes in the MIT-BIH dataset. The network had four 1D-convolution and max pooling layers, followed by two ReLU layers, and a fully connected softmax layer to classify beats into five classes. The model was tested by randomly splitting the beats into training and testing sets, and got 98.3% accuracy.

In this paper, we build on the previous discussion and use the existing CNN and LSTM models as a basis for our design. Our aim is to improve these models and achieve better performance in ECG. classification. We compare different architectures and hyperparameters and evaluate their performance on the MIT-BIH dataset [6]. The accuracy and other metrics for each one of them have been reported and used to identify which models are the best for this problem.

Following on from work discussed in this section, the first CNN model that we proposed for arrhythmia classification is shown in Figure 4. The model consists of two 1D convolutional layers with 64 and 32 filters, with kernel size of 16 and 8 accordingly. Each CNN layer is followed by a batch normalization layer and a max pooling layer with a pool size of 2. The output is flattened and passed through four dense layers with 512, 128, 32, and 9 neurons. The first three dense layers use the ReLU activation and the last layer uses SoftMax activation. We use Adam optimizer, categorical cross entropy loss, and accuracy metric to train the model with a batch size of 64 for 5 epochs.

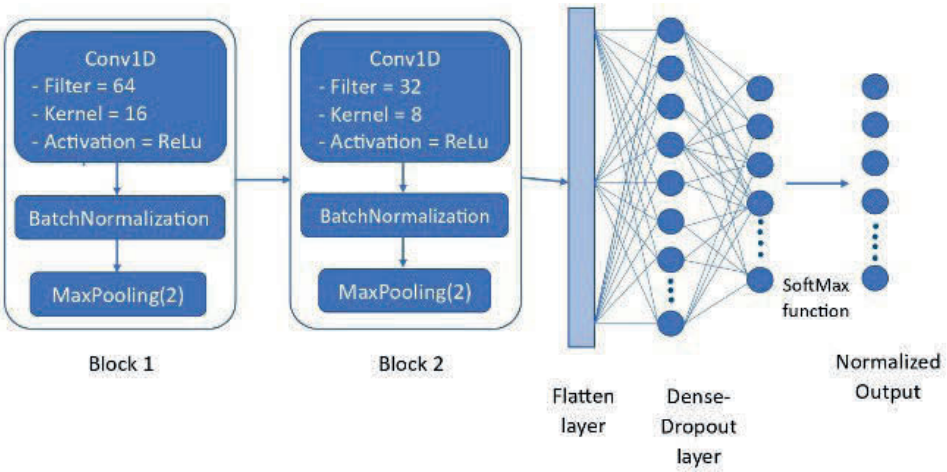


Figure 4: Proposed 1D CNN model architecture.

The second CNN model that we used consists of six 1D convolutional layers. Each convolutional layer applies a one-dimensional filter to the input and uses ReLU activation and 'same' padding. The number of filters increases from 64 to 256 as the layers go deeper, which means that the model can learn more complex and abstract features. Each convolutional layer is followed by a batch normalization layer, which normalizes the output and improves the training speed and stability. After every two convolutional layers, there is a max pooling layer, which reduces the dimensionality of the output by taking the maximum value in each window of size two. There is also a dropout layer, which randomly sets a fraction of input units to zero during training, which helps prevent overfitting. A flatten layer is used to convert the multidimensional sequences into a one-dimensional vector. Then four dense layers are added, the first three dense layers have 512, 128, and 32 neurons, respectively, and use ReLU activation and are followed by batch normalization and dropout layers. The last dense layer has nine neurons and uses SoftMax activation, which is suitable for multi-class classification tasks.

For the third CNN model we used three 1D convolutional layers with 32, 64, and 128 filters, each followed by batch normalization and max pooling layers with pool size of 2. The output is flattened and passed through dense layers with 256, 64, and 9 neurons. After the first two dense layers there is a dropout layer with a rate of 0.2 and they use ReLU activation. The last dense layer uses SoftMax activation.

All the CNN models that we experimented with in this paper have the same training settings. They use Adam optimizer, categorical cross entropy loss, and the models are trained with a batch size of 256 for 5 epochs [8].

The first LSTM model proposed and investigated in this paper uses the Keras Sequential API and consists of 8 weighted layers. The first layer has 128 units and returns sequences. After the first LSTM layer a dropout regularization with a rate of 0.2 is applied. Dropout randomly sets a fraction of input units to 0 during training, which helps prevent overfitting. The second LSTM layer has 64 units and after this layer a dropout regularization with a rate of 0.2 is applied. Next to reduce the temporal dimensions of the sequences a Max-Pooling layer with a pool size of 2 is applied. Then a flatten layer is used to convert the multidimensional sequences into a one-dimensional vector, preparing the data for the fully connected layers. The model uses three dense layers, the first dense layer has 256 neurons and uses the ReLU activation function, the second dense layer has 128 neurons and uses the same activation function ReLU. After each of these two dense layers a dropout regularization with a rate of 0.3 is applied. The final dense layer has 9 neurons with the SoftMax activation function, suitable for multi-class classification tasks.

The second LSTM model has two LSTM layers with 256 and 128 units, each followed by a dropout layer with a rate of 0.3. Then a dense layer with 64 neurons with a ReLU activation is applied, followed by a dropout layer with a rate of 0.3 and another dense layer with 9 neurons with SoftMax activation function.

The third LSTM model is similar to the second, but instead of two LSTM layers it has three LSTM layers with 512, 256 and 128 units. We increased the dropout rates for each dropout layer in the third model to 0.5 which are higher than the 0.3 of the second model, to increase the robustness and generalization ability of the model and also to prevent overfitting.

All LSTM models use the Adam optimized, with the default learning rate of 0.001, categorical cross-entropy loss and accuracy metric. The models are fit on the training dataset for 5 epochs and 256 batch size [9].

For the remaining models, the same hyper-parameters as in [3] were used: GBC and ADA: n_estimators = 100; RFC: n_estimator = 10 and max_depth = 10 ; SVC: default parameters.

The proposed models were trained on a HP notebook equipped with an Intel Core i5-7200U with 2 cores, running at 2.50GHz (2.71GHz turbo boost). 8GB internal RAM and 1TB internal SSD hard drive. Substantial computational power and training time were needed to train the CNN and LSTM models. The SSD storage enabled fast data access, and the setup was affordable. While the absence of the GPU made the training times longer, but still the project was completed successfully on this hardware setup.

5 RESULTS

A comprehensive representation of the performance metrics of the ten classification models used in this paper is provided in Figure 5. Specifically, it displays the accuracy and standard deviation for each of these models, allowing for a visual comparison of their predictive capabilities and consistency. These results indicate that CNN is the best model for arrhythmia detection, as it correctly classified most of the ECG beats in all classes and made fewer errors than the other models. However, LSTM and SVC also performed well and may have some advantages over

CNN in terms of computational complexity and robustness, respectively.

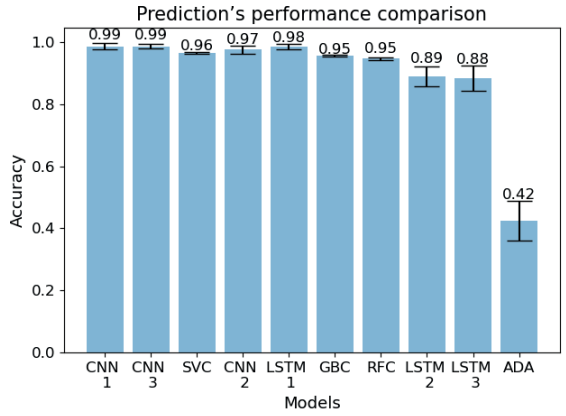


Figure 5: Achieved accuracy for each of the 10 models.

To measure and compare the performance of the ML models and estimate its general performance, cross-validation is used. The data is split into 75% of training set and 25% of test set. The accuracy and other metrics are reported on the test set as the performance of the classifiers. To have a clear view of the model's performance on each class, we measured three metrics: precision, recall, and F1-score, for each of the classes. The results also show the weighted average of these metrics across all classes, taking into account the imbalance of the classes. Due to conciseness and in order to highlight the most relevant findings, in Figure 6 we will showcase results for the top three best-performing models only.

The classification results show that the three CNN models had the highest accuracy of around 95% to 96%, followed by the SVC model with 94% and the first LSTM model with 93% accuracy. The first graph in Figure 6 illustrates the results achieved using the first CNN classifier. It is obvious that the CNN model succeeded in predicting all the classes with different performances. The most predictable classes for CNN are the N and '/' classes, the highest recall (100%) is achieved for the L class and F1-score (99%) for the R class. The second graph show the results achieved by the SVC model, which has similar results as the CNN model, the highest recall (99%) and F1-score (99%) for the L and R classes. The best LSTM model of the three LSTM models had the highest recall (100%) and F1-score (99%) for the L class. The other models achieved significantly lower results, i.e., GBC: 91%, RFC: 90%, ADA: 26%. This shows that the ADA classifier significantly underperforms with the default hyperparameters, requires hyperparameter tunning.

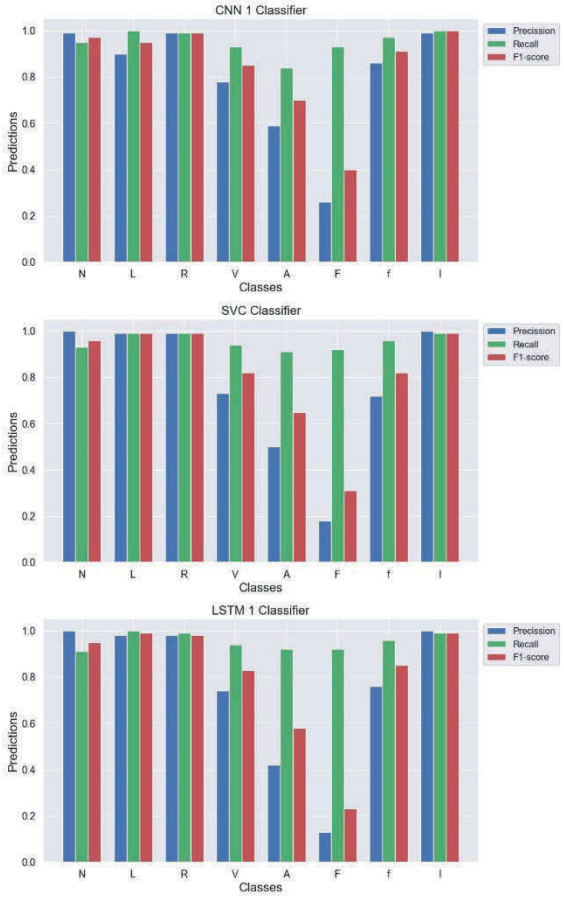


Figure 6: Graphical representation of the classification report for the top 3 models according to accuracy.

Figure 7 illustrates the confusion matrices for the three best models, the diagonal of these confusion matrices shows that most of the models have more than 80% accuracy in classifying the ECG beats. The L, R and / classes are the easiest to classify, with over 90% accuracy for all the models. The N class is the hardest to classify, especially for ADA, which only has about 20% accuracy. The other models have over 80% accuracy for the N class. ADA is very sensitive to noise and outliers, and it learns gradually. Therefore, ADA is not suitable for ECG beat classification. It can be noticed that classes A and F are the most difficult to classify correctly. This might be explained by the fact that class A beats have a similar shape to class N beats, which makes them hard to distinguish. On the other hand, class F beats are very uncommon in the dataset, and the simple method of up-sampling them does not help the models to learn their features well enough.

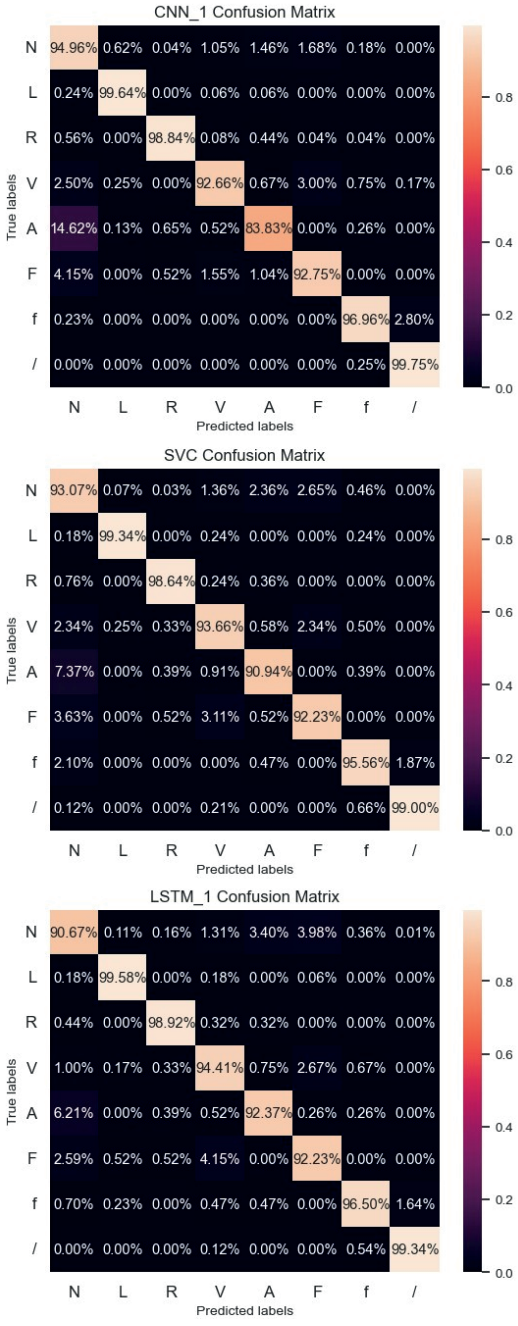


Figure 7: Confusion matrices for the top 3 models according to accuracy.

6 CONCLUSIONS

The paper presented a thorough comparison of 2 Deep Learning approaches (CNN and LSTM) to 4 classical Machine Learning models (GDC, ADA, RFC, SVC) to classify arrythmia from ECG data. In total we used 10 models (3 for each DL approach and 1 for each ML approach) and compared their performance on the MIT-BIH Arrhythmia database. Various metrics such as accuracy, precision, recall and F1-score we used to evaluate the models on different classes od ECG beats. The results showed that CNN is the best model overall for arrhythmia detection, achieving the highest accuracy of 95% and the highest F1-score for most of the classes. SVC and LSTM also performed well, with accuracy of 94% and 93%, respectively, and high F1-scores for some classes. However, LSTM and SVC may have some advantages over CNN in terms of computational complexity and robustness, respectively. The ECG data can present several challenges for the arrhythmia detection models, such as noise, imbalance, and similarity of classes. To overcome these challenges, more advanced preprocessing and resampling techniques should be used. Incorporating domain knowledge and expert feedback into the models is a promising direction for future research and development. We believe that with further research on improving the performance and interpretability of machine learning models for ECG data analysis, more precise results can be expected.

REFERENCES

[1] K.Mc Namara, H. Alzubaidi, and J.K. Jackson, “Cardiovascular disease as a leading cause of death: how are pharmacists getting involved?” Integrated Pharmacy Research and Practice, vol. Volume 8, no. 8, pp. 1-11, Feb. 2019, [Online]. Available: https://doi.org/10.2147/iprp.s133088.

[2] A. Ullah, S.M. Anwar, M. Bilal, and R.M. Mehmood, “Classification of Arrhythmia by Using Deep Learning with 2-D ECG Spectral Image Representation,” Remote Sensing, vol. 12, no. 10, p. 1685, May 2020, [Online]. Available: https://doi.org/10.3390/rs12101685.

[3] S. Verma, “Development of Interpretable Machine Learning Models to Detect Arrhythmia based on ECG Data,” May 2022, [Online]. Available: https://doi.org/10.48550/arxiv.2205.02803.

[4] S.K. Pandey and R.R. Janghel, “Automatic detection of arrhythmia from imbalanced ECG database using CNN model with SMOTE,” Australasian Physical & Engineering Sciences in Medicine, vol. 42, no. 4, pp. 1129-1139, Nov. 2019, [Online]. Available: https://doi.org/10.1007/s13246-019-00815-9.

[5] M.T. Le, V.S. Rathour, Q.S. Truong, Q. Mai, P. Brijesh, and N. Le, “Multi-module Recurrent Convolutional Neural Network with Transformer Encoder for ECG Arrhythmia Classification,” Jul. 2021, [Online]. Available: https://doi.org/10.1109/bhi50953.2021.9508527.

[6] S. Armstrong, “Survey of Machine Learning Techniques To Predict Heartbeat Arrhythmias,” arXiv.org, Aug. 22, 2022, [Online]. Available: https://arxiv.org/abs/2208.10463, [Accessed on Nov. 04, 2023].

[7] J. Wu et al., “Editor’s Choice - Impact of initial hospital diagnosis on mortality for acute myocardial infarction: A national cohort study,” European Heart Journal: Acute Cardiovascular Care, vol. 7, no. 2, pp. 139-148, Aug. 2016, [Online]. Available: https://doi.org/10.1177/2048872616661693.

[8] S.K. Pandey and R.R. Janghel, “Automatic detection of arrhythmia from imbalanced ECG database using CNN model with SMOTE,” Australasian Physical & Engineering Sciences in Medicine, vol. 42 (4), pp. 1129-1139, November 2019.

[9] J. Gao, H. Zhang, P. Lu, and Z. Wang, “An Effective LSTM Recurrent Network to Detect Arrhythmia on Imbalanced ECG Dataset,” Journal of Healthcare Engineering, October 13, 2019.

[10] A. Mustaqeem, S.M. Anwar, and M. Majid, "Multiclass classification of cardiac arrhythmia using improved feature selection and SVM invariants," Computational and Mathematical Methods in Medicine, vol. 2018, pp. 1-10, 2018, doi: 10.1155/2018/7310496.

[11] A. Mustaqeem, S.M. Anwar, M. Majid, and A.R. Khan, “Wrapper method for feature selection to classify cardiac arrhythmia,” IEEE Xplore, Jul. 01, 2017, [Online]. Available: https://ieeexplore.ieee.org/document/8037650, [Accessed on Jan. 11, 2023].

[12] G.B. Moody and R.G. Mark, “The impact of the MIT-BIH Arrhythmia Database,” IEEE Engineering in Medicine and Biology Magazine, vol. 20, no. 3, pp. 45-50, 2001, [Online]. Available: https://doi.org/10.1109/51.932724.

Mitigating the Ronin Protocol Vulnerability in the Context of RBAC Policy

Halyna Huzenko and Leonid Galchynsky
Educational and Research Institute of Physics and Technology, Igor Sikorsky Kyiv Polytechnic Institute, Beresteiskyi Avenue 37, Kyiv, Ukraine
gguzenko989@gmail.com, hleonid@gmail.com

Keywords: Vulnarability, Ronin Protocol, RBAC Standard, ERD, Simulation Modelling.

Abstract: The article discusses the structure of the Ronin protocol and its components, focusing on consensus mechanisms and validators. The purpose of the study was to identify the vulnerability of the protocol and to develop methods for its resolution. It was determined that the bridge component of the protocol has a certain vulnerability. Analyzing and investigating the structure and mechanics of Ronin smart contracts, it was found that all validators are Bridge Validators. This prompted a more detailed study of the protocol structure. Audits for 2022 and 2023 were analyzed, which indicated the presence of privileged functionality in some roles in the system. The conclusion was that the protocol has an unformalized role-based access distribution model. By comparing with the NIST standard, it was found that the role-based access control system in the Ronin protocol (Ronin RBAC) is a Flat Model. By increasing the level of the model to the level of the Restricted Model, it was possible to increase the security level of the protocol. Using the MySQL environment, a simulation model was developed that confirmed the vulnerability of the considered access control system. Based on the analysis of the standard requirements, steps were formulated to make changes to the simulation model. To solve this problem, it was proposed to change the role model of access distribution to Level 3 of the NIST RBAC standard.

1 INTRODUCTION

Decentralized finance is becoming increasingly popular. DeFi protocols that use blockchain technology for secure and transparent record-keeping are becoming increasingly widespread. The key area of blockchain application is cryptocurrencies, and a notable innovation is sidechains.

A sidechain is a type of blockchain scaling solution that allows the creation of new independent blockchain networks that can interact with an existing blockchain. They are considered to be the 2nd layer of the Ethereum protocol, extending and inheriting its security guarantees while providing scalability and speed. The relevance lies in the fact that they are able to eliminate some of the scalability and functionality limitations of existing blockchain networks such as Bitcoin and Ethereum.

This allows for the creation of specialized ones that can help reduce the load on the main network, increase throughput, and improve overall performance.

The simple blockchain architecture hides a serious problem: it is impossible to achieve decentralization, security, and scalability at the same time. So, to get a secure and decentralized blockchain, you need to sacrifice scalability.

This problem can be solved with the help of sidechains - independent blockchains with their own security rules. A layer 2 blockchain interacts with Ethereum by sending transaction packets to ensure security and decentralization without changing the Ethereum protocol. This allows layer 1 to manage data security and availability, while layer 2 provides scalability. Layer 2 sends completed proofs back to layer 1 and removes the transaction load from layer 1. Many sidechains also include binding mechanisms to securely move assets between the main blockchain network and the sidechain.

One of the well-known sidechain implementations is the Ronin protocol, developed specifically for the Axie Infinity game [1]. Initially, it used the Proof-of-Authority consensus algorithm for low fees and fast transactions, but later switched to a hybrid of Proof-of-Authority and Proof-of-Stake

mechanisms. Ronin's main goal is to increase transaction efficiency and profitability, serving as a valuable solution for blockchain-based games, especially in the Play-to-Earn (P2E) genre. Developers are increasingly exploring its adaptability for other games in the blockchain ecosystem.

However, while solving the problem of the underlying network to some extent, the Ronin protocol has its own problems, primarily security issues.

We have organized this paper as follows: in Section 2, we describe the Ronin protocol and analyze its vulnerabilities. In Section 3, we present the result: a new RBAC model of the Ronin protocol and conduct simulations in the MySQL Workbench environment to show how the vulnerability of the Ronin protocol can be reduced by changing the roles of network participants.

In the fourth section, we summarize the work and offer suggestions for further research.

2 STRUCTURE OF THE RONIN PROTOCOL AND ITS VULNERABILITIES

The Ronin protocol is developed by Axie Infinity, one of the leading blockchain game developers. The goal is to introduce a traditional economic system into the P2E (play-to-earn) game model.

2.1 Structure of the Ronin Protocol

The main elements of the sidechain are:

1) The main blockchain: this is the blockchain used to store the main data and ensure consensus between users.
2) Smart contracts: These are applications that enable complex transactions on the blockchain, including the transfer and storage of assets. Smart contracts can be deployed both on the main blockchain and on sidechains.
3) Sidechain is a blockchain that runs in parallel with the main blockchain, providing security and consensus with it. Sidechains can have their own consensus rules and security mechanisms.
4) Bridges: these are elements that allow assets to be transferred between sidechains and the main blockchain. Bridges provide security and consensus between blockchains.
5) Messaging protocols: These are protocols that allow messages to be exchanged between sidechains and the main blockchain [2].

Transactions can be used to transfer assets between users, make payments within the network, and interact with third-party services such as network researchers, exchanges, etc.

Smart contracts are used to define the conditions for executing transactions, where the contract code is executed automatically when certain conditions are met [3].

The Ronin protocol supports all tokens, including the main regulatory token RON. Users can pay for transactions on the platform and use DeFi features such as community management and rewards through validators. With the PoS (Proof of Stake) mechanism, validators stake 250,000 RON to receive the right to verify blocks and reward in RON. RON holders can delegate their tokens and become delegates of validators, receiving rewards in RON and participating in network governance. The rewards are distributed between validators and their delegates, and if a validator fails to create correct blocks, its share of RON may decrease, as well as the public share of RON for accepting applications [4].

2.2 Ronin Consensus Algorithm

The Ronin protocol uses a PoA (Proof of Authority) network as a verification mechanism. In the Ronin network, authorities are certain nodes that are responsible for confirming transactions and creating new blocks. They are pre-selected by Sky Mavis itself. Each authority has its own public key, which it uses to sign blocks and transactions. The advantage of PoA is that it does not require the high computing power and energy consumption required for PoW (Proof of Work) mining. In addition, PoA provides speed and scalability of transactions, which allows Ronin to process large amounts of data. However, PoA also has its drawbacks: it can be less decentralized than PoW, as control over the network is in the hands of elected authorities. In addition, PoA does not allow users to mine and receive rewards for maintaining the network. This algorithm is too centralized and has led to the possibility of the largest attacks [5].

To modernize the consensus mechanism, the developers chose the modifiable DPoS. Its advantage over PoS is the presence of delegates, which makes the network even more decentralized [6]. The difference between these two mechanisms can be seen in Figure 1.

The most significant change is that token holders can vote for themselves or delegate a share to a representative. The more tokens a validator receives, the higher his or her chance of being elected. This is

how the validator reliability system works. Each block created on the network must be confirmed by validators, otherwise it can be fined. This helps to maintain network security and the discipline of validators. Thus, the interaction between validators and the Ronin protocol ensures the security, stability, and efficiency of the network.

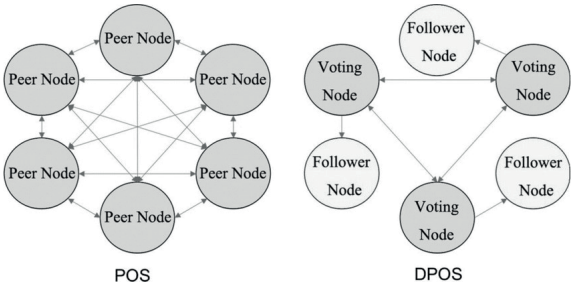


Figure 1: The difference between consensus mechanisms [7].

2.3 Smart Contracts of Ronin Protocol

Like other protocols for blockchain networks [8, 9], the RONIN protocol has a list of smart contracts. By definition, a smart contract is a program code that contains terms and rules that are automatically executed under certain conditions. Smart contracts are a safe and reliable means of ensuring compliance with the terms of contracts, as they are executed automatically without the involvement of intermediaries. Currently, there is a package of smart contracts that support the fulfillment of DPoS consensus conditions. Smart contracts in the Ronin protocol have an impressive set of features: In particular,

1) Transaction processing: Smart contracts in RON can be used to automatically process transactions between users.
2) Token creation: Smart contracts can be used to create and manage tokens on RON. This can be useful for issuing your own tokens, stablecoins, or organizing an ICO.
3) Revenue distribution: Smart contracts can be used to automatically distribute revenue to different participants.
4) Organization of voting: Smart contracts can be used to organize voting for various decisions. The contract can collect votes from different participants and automatically tally the results.
5) Asset management: Smart contracts can be used to manage assets in the marketplace. For example, the contract can automatically distribute profits.

According to the audit by Verichains Lab, the interaction of smart contracts and sidechains with the result of recording the main information looks like in Figure 2:

It is also worth noting such an important type of smart contract as the Ronin Bridge smart contract. Ethereum and Ronin are two separate networks with different protocols, and assets cannot be directly transferred between them. To transfer assets from Ethereum to Ronin or vice versa, users must go through the Ronin Bridge. Ronin Bridge smart contracts are written in Solidity using the OpenZeppelin library and are designed to support multiple chains. When an event deposit occurs on the main chain, the Bridge validator component picks it up and passes the corresponding transaction to Ronin. For output and control events, it starts with Ronin and then passes it to the other chains.

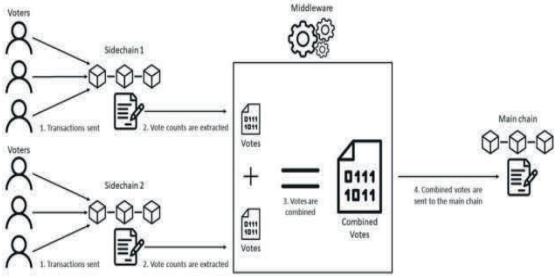


Figure 2: Scheme of interaction between smart contracts and sidechains [10].

2.4 Vulnerability Analysis of the Ronin Protocol

Like all blockchain networks, Ronin is at risk of hacker attacks, as these peer-to-peer exchanges, unfortunately, have vulnerabilities. The topic of vulnerability of blockchain systems has recently received considerable attention, with in-depth analytical studies providing a taxonomy of vulnerabilities that correspond to the general blockchain technology stack [11]. However, there is a lack of such materials specifically for the Ronin network. The fact that the issue of Ronin vulnerabilities is currently relevant is evidenced by the fact that in March 2022, the Ronin network suffered one of the largest DeFi hacks to date. The attackers stole approximately USD 624 million. The hack took place through compromised Ronin and Axie DAO validator nodes. This resulted in the compromised private keys being used to fake two fake network hijackings that swallowed hundreds of millions of cryptocurrency from the bridge. An investigation by independent companies found that

the Ronin protocol was too centralised, with no clear separation of access and monitoring. At least one of the validator nodes had privileged permission to provide its signature to other nodes via a whitelist. The CryptoPotato audit [12] raised the issue of the PoA mechanism's flaw and what causes it - the consensus on the proof-of-authority rule, which is part of the protocol's problems. The audit focused on smart contracts, code and database quality, compliance of smart contract logic with requirements, cross-references to the structure, and implementations based on similar smart contracts. The analysis showed that the "significant" category includes functions with the problem of privilege distribution in the role-based management model. The "significant" category indicates that the built-in logic of rights and privileges is quite vulnerable and requires appropriate changes.

As an example, for the MainChain-GovernaceAdmin smart contract, the Governor role can be assigned the Relayer_Role, which will elevate it to GovernanceRelay. And this, in turn, can lead to critical roles being assigned to validators, allowing them to administer the blockchain network. This is unacceptable, as, by definition, a validator is a cryptocurrency user who verifies blocks and transactions, and only that. Granting an administrator role to a validator account is an abuse of privilege. Obviously, the Ronin RBAC protocol has an informal model of access delimitation, which is manifested in the available roles and the provision of access to certain functionality only if the role is available. In particular, assigning an administrative role to a validator leads to a conflict of interest when a hacker, having hacked into the validator's account, can immediately gain administrator access. We need to look for ways to solve this problem. And using the NIST RBAC role assignment model [13] can help.

3 RESULTS

Role-Based Access Control (RBAC) is a security methodology that is based on managing user access to protect resources and is a standard developed by the National Institute of Standards and Technology (NIST) of the United States. Currently, according to this standard, the separation of roles can be implemented at one of four levels, each of which has the capabilities of the previous one:

1) Flat RBAC.
2) Hierarchical RBAC.
3) Limited RBAC.
4) Symmetric RBAC.

Given the limited size of the article and the needs of use in building the model, we will limit ourselves to considering the first three levels, which are sufficient for model development.

Flat RBAC can support user access permissions to resources through roles. In this case, the procedure for granting user roles on a many-to-many basis, as well as assigning many-to-many permissions to roles. Users can use the permissions of several roles at the same time.

Hierarchical RBAC supports all the functionality of flat RBAC and additionally supports a hierarchy of roles (partial, arbitrary, or limited order).

Restricted RBAC has all the functionality of hierarchical RBAC and must additionally provide a separation of duties (SOD). Figure 3 shows a schematic representation of hierarchical RBAC with restrictions.

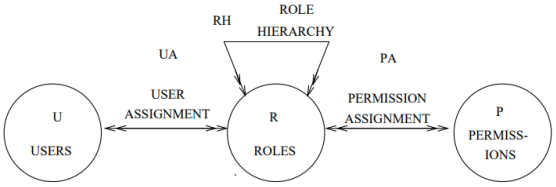


Figure 3: Hierarchical RBAC with restrictions. [16].

3.1 The RBAC Ronin Model

Based on the protocol description, we define roles and resources as entities as follows:

1) Governor(ID, ProposalDurance, Trusted Organisation, Signature) - responsible for the validators that were selected by the Proof of Authority (PoA) consensus mechanism. This entity has access to vote on proposals and submit them.
2) Validator(ID, OnlyAdminID, GovernorID, RelayerRoleID) - an entity representing a regular validator who is assigned certain roles. Each validator has a unique identifier (ID) and can have additional roles, such as the administrator role (OnlyAdminID), the governor role (GovernorID), and the relayer role (RelayerRoleID) to transmit offers from another network.
3) RelayerRole(ID) - an entity that has the right to relay offers from another network.
4) OnlyAdmin(ID, Access) - an entity representing a regular validator to whom certain roles are assigned. Each validator has a unique identifier (ID) and can have additional roles, such as the administrator role (OnlyAdminID), the governor role (GovernorID), and the relayer role

(RelayerRoleID) to forward offers from another network.

According to the description of the protocol, the relationships between entities can be shown in the Figure 4 as an ERD diagram in UML notation. Let's see what level of the RBAC standard the model shows

1) Ronin users access functions through rolesю. For example, only onlyAdmin gets permission to change thresholds.
2) Ronin provides the ability for one user to assume multiple roles, while the same.
3) Roles in Ronin can be assigned to many permissions.
4) The logic of the protocol states that the Validator can use the Validator and GovernorID permissions.
5) All four of these features correspond to the flat RBAC level.

Since there are no hierarchical relationships and restrictions in the scheme, it can be argued that the Ronin protocol implements the first level according to the NIST standard, namely flat RBAC.

3.2 Modeling the Vulnerability of the Ronin Network

Let's now show the existence of a Ronin protocol vulnerability in this role model by simulating it. We will use MySQL Workbench as a simulation environment. Roles that will be further considered and implemented:

OnlyAdmin, RELAYER_ROLE, and OnlyGovernor. These roles can be assigned to the validator user, which is the user who appeared in the previous sections and attack scenarios. For modeling, a database corresponding to the ERD in Figure 4 was developed and several procedures were created: PauseContract; ResumeContract; UpdateContract-Address; SetThreshold; RelayProposal; Governor-Propose; ReplaceBridgeValidator; GovernorVotes; ChangeTrustedOrganization.

Creating the Validator entity looks like this:

```
CREATE TABLE IF NOT EXISTS `Validator`
(
  `ID` int NOT NULL PRIMARY KEY,
  `OnlyAdminID` INT, `GovernorID` INT,
  `RelayerRoleID` INT,
  FOREIGN KEY (`GovernorID`) REFERENCES
`onlyGovernor`(`ID FOREIGN KEY
(`RelayerRoleID`) REFERENCES
RelayerRole`(`ID
  FOREIGN KEY (`OnlyAdminID`)
REFERENCES `onlyAdmin`(`ID`).
```

This means that a validator can have several roles if they are assigned by the corresponding smart contract:

```
INSERT INTO `Validator` (`ID`,
`OnlyAdminID`, `GovernorID`,
`RelayerRoleID`)
VALUES
(1, 87250316, 91263457, 38529741)
```

– has all the roles,

`(2, NULL, 57389126, NULL)` – validator is assigned via PoA,

`(3, NULL, NULL, NULL)` – the validator is assigned via DpoS.

After describing all three validators with the DESCRIBE statement, we will execute several test queries to confirm at least one attack scenario in which a validator tries to perform the administrative function of changing the contract address. On Figure 5 validators who have not been assigned this role will be denied access.

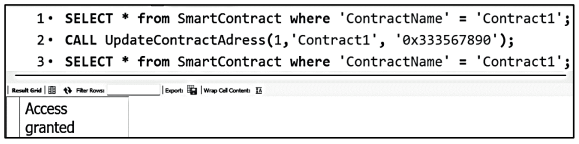


Figure 5: Access of Roles.

A validator with the Administrator role assigned, even temporarily, has access to critical functions, including changing the OnlyGovernor Trusted Organization. After that, the hacker can access the validator's signature through a fictitious organization.

The hacker can also change the delay time between smart contracts to execute them at the same time. Then, returning all the Trusted Organizations with the Validators' signatures. Then the hacker is able to take over the network as follows:

1) Obtain 10 out of 22 validator positions in the Delegated Proof of Stake (DPoS) mechanism.
2) Intercept the session information of a validator wallet that has been recently spotted with administrator activity by capturing cookies.
3) Obtain a signature and take over the account.

We will not describe further steps, but cyber incidents with the Ronin network began with this.

How can this vulnerability be mitigated? Our answer is to change the role policy.

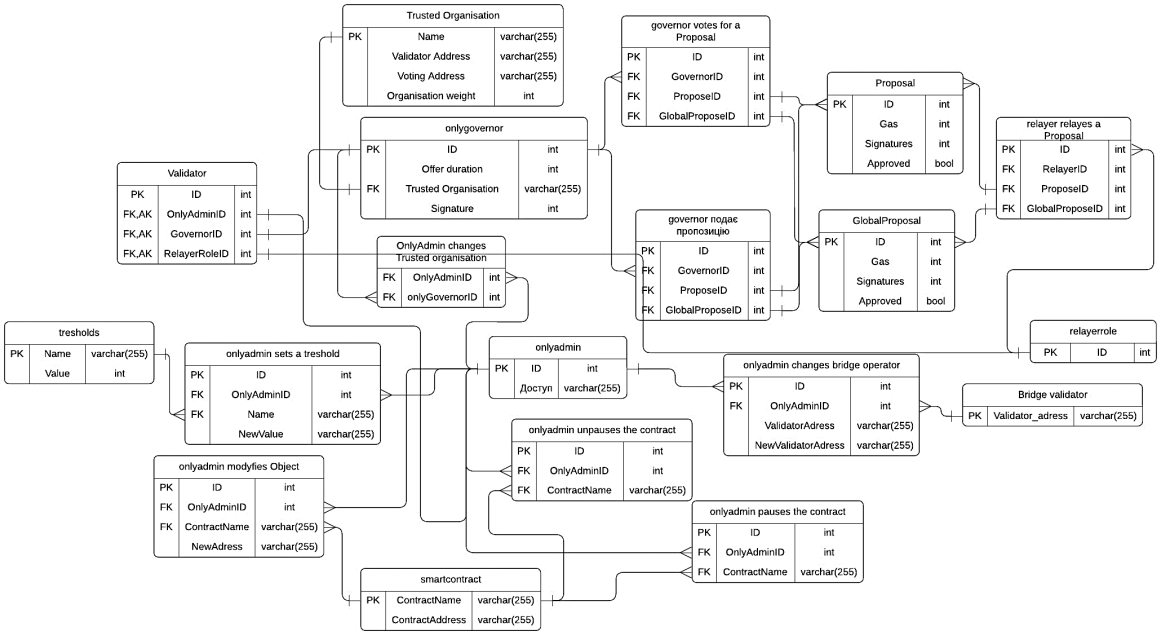


Figure 4: ERD diagram of the Ronin protocol in UML notation.

3.3 Change of Roles in the Ronin Protocol

Let's try to raise Ronin's RBAC level to higher levels. To avoid accidentally assigning the RelayerRole to a user with the onlyGovernor role, which could lead to an unauthorized increase in privileges, we will make the RelayerRole a parent and the onlyGovernor a child. Then the system will always see which of the validators has the RelayerRole. At the same time, the OnlyAdmin role should be made the highest, that is, in our case, parent to the RelayerRole. Thus, the Ronin role model is raised to the level of hierarchical RBAC.

The Validator mustn't have access to administrative functions, because its address is in the public network and it is more vulnerable. This can only be realized through restriction. It is necessary to restrict the role of OnlyAdmin to use the functionality by the separation of duties (SOD), and it will be appropriate to use a static restriction so that the user who has the Validator client on his system does not have access to administrative functions in case the attacker left a backdoor. This will block many potential attack vectors. Thus, the protocol will have RBAC of the third level according to the NIST standard.

The RelayerRole hierarchy was added using the GovernorID foreign key:

```
CREATE TABLE IF NOT EXISTS
`RelayerRole` (
  `ID` int NOT NULL PRIMARY KEY,
  `GovernorID` INT UNIQUE NOT NULL,
   FOREIGN KEY (`GovernorID`)
REFERENCES `onlyGovernor`(`ID`)
Conclusions
```

To implement a static restriction, you need to create a trigger to prohibit a user from having the OnlyAdmin role with any other role.

```
DELIMITER //
CREATE TRIGGER forbid_onlyadmin
FOR EACH ROW
BEGIN
    DECLARE governor_count INT;
    DECLARE admin_count INT;
    DECLARE relayer_count INT;
         SELECT COUNT(*) INTO
governor_count FROM `onlyGovernor`
WHERE `ID` =
NEW.`GovernorID`;
    SELECT COUNT(*) INTO admin_count
FROM `onlyAdmin` WHERE `ID` =
NEW.`OnlyAdminID`;
    SELECT COUNT(*) INTO relayer_count
FROM `RelayerRole` WHERE `ID` =
NEW.`RelayerRoleID`;
       IF governor_count > 0 AND
admin_count > 0 THEN
        SIGNAL SQLSTATE '45000'
```

```
            SET MESSAGE_TEXT = 'Cannot
have both OnlyGovernor and OnlyAdmin
roles';
  END IF;
 IF governor_count > 0 AND
relayer_count > 0 THEN
        SIGNAL SQLSTATE '45000'
            SET MESSAGE_TEXT = 'Cannot
have both RelayerRole and OnlyAdmin
roles';
...
```

The above role changes entail changes in the ERD for the Ronin protocol. Unfortunately, the length of the article does not allow us to provide an updated diagram, but we note that the Validator has changed to a User, which will not allow the Validator to get the Administrator role. The RelayerRole is now a parent to OnlyGovernor, which will allow you to assign individual validators, while simplifying monitoring and management.

Now let's show the same attack scenario for the modified ERD, i.e., we will make queries to the updated database. We will use the same roles as before the ERD change.

When we try to assign all the roles to the user, we get an error (Figure 6).

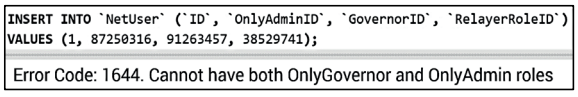


Figure 6: Trigger for unauthorised role assignment.

When we try to assign OnlyAdmin and another role to the user, we also get an error (Figure 7).

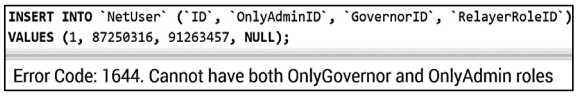


Figure 7: Trigger for unauthorised role assignment.

Therefore, it can be argued that the vulnerability that allowed hackers to take over the network can be eliminated by changing the role policy of the Ronin protocol.

4 CONCLUSIONS

The relevance of identifying and eliminating Ronin protocol vulnerabilities is caused by high-profile cyber incidents with thefts of hundreds of millions of dollars. A detailed analysis shows that the Ronin protocol has numerous vulnerabilities that lead to such incidents. This is especially true for the vulnerabilities related to the consensus mechanism, in which validators were given unjustifiably broad rights, which led to the vulnerability. This vulnerability allowed hackers to take over the network, resulting in multimillion-dollar losses for the network participants. The vulnerability was rooted in a flawed role-based rights policy. A simulation model of the Ronin protocol was developed, which showed the possibilities of illegal penetration under the existing role-based policy, which can be qualified as a Flat Model of the NIST RBAC standard. It was shown that changing the role policy to the level of the Restricted Model of the NIST RBAC standard eliminates the consensus vulnerability, which is currently the main source of user losses. The obtained result shows that the applied methodology gives the prospect of continuing research to combat other attack vectors on the Ronin network.

REFERENCES

[1] M. S. Mavis, "Official Ronin Whitepaper: Consensus," Apr. 28, 2023.

[2] M. S. Mavis, "Official Axie Infinity Whitepaper," Jan. 1, 2023.

[3] M. Bartoletti, S. Carta, T. Cimoli, and R. Saia, "Dissecting Ponzi schemes on Ethereum: Identification, analysis, and impact," Future Generation Computer Systems, vol. 102, pp. 905-913, Aug. 2019.

[4] E. Castronova et al., "As real as real? Macroeconomic behavior in a large-scale virtual world," New Media & Society, vol. 11, no. 5, pp. 685-707, 2009.

[5] R. Behnke, "Explained: The Ronin Hack," Mar. 30, 2022.

[6] V. B. Vishal and A. B. Aniruddha, "Preferential Delegated Proof of Stake (PDPoS) – Modified DPoS with Two Layers towards Scalability and Higher TPS," 2023.

[7] S. Wan et al., "Recent advances in consensus protocols for blockchain: A survey," Springer Science+Business Media, LLC.

[8] M. Alharby et al., "Blockchain-based smart contracts: A systematic mapping study of academic research," in 2018 ICCBB, IEEE, pp. 1–6, 2018.

[9] S. N. Khan et al., "Blockchain smart contracts: Applications, challenges, and future trends," Peer-to-Peer Netw. Appl., vol. 14, pp. 2901–2925, 2021.

[10] Verichains Lab, "Report for Sky Mavis: Security Audit – Ronin Bridge Smart Contracts. 1.1 - Public Report," Jun. 28, 2022.

[11] X. Li et al., "A survey on the security of blockchain systems," Future Generation Computer Systems, pp. 841–853, 2020.

[12] J. Lyanchev, "The Biggest Ever Crypto Hack: What Happened in the Ronin Bridge Attack on 'cryptopotato'," Mar. 30, 2022.

[13] D. Ferraiolo and R. Kuhn, "Role-Based Access Controls: Conference 15th National Computer Security Conference (NCSC)," Oct. 13-16, 1992.

Analysis and Implementation of an Efficient Traffic Sign Recognition Based on YOLO and SIFT for Turtlebot3 Robot

Stefan Twieg and Ravin Menghani
Department of Electrical, Mechanical and Industrial Engineering, Anhalt University of Applied Sciences, Bernburger Str. 55, Köthen, Germany
stefan.twieg@hs-anhalt.de, ravinvijaybhai.menghani@student.hs-anhalt.de

Keywords: Traffic Sign Recognition, Machine Learning, YOLO, SIFT, Turtlebot3, ROS, Robot Operating System, Convolutional-Neural-Network, CNN.

Abstract: Traffic Sign Recognition (TSR) is one of the key aspects for autonomous driving and it plays a vital role to make autonomous driving successful, but that's only possible if TSR is efficient enough and reliable. This work addresses exploration of simple and fast to implement options for robotic applications. For analysis and implementation, we are focusing on a Turtlebot3 Robot (TB3). Various potential TSR algorithms are evaluated in different test-cases with the goal of developing an optimized TSR with accurate results for German traffic signs. Therefore, the robot was tested on its own Mini-City track. On this Track we started to detect the signs with a simple Scale-Invariant Feature Transform (SIFT). However, the accuracy of SIFT was showing limitations for the use within TSR on mini-city-Track. This approach focuses on educational use where limitations and simple applications of autonomous driving are investigated. A review of state-of-art algorithms was done, to evaluate and improve accuracy. For example, Oriented FAST and Rotated Brief algorithm (ORB), You Only Look Once (YOLO) and SIFT algorithm was tested on TB3 in a way that all important criteria are fulfilled along with system being real-time. Regarding YOLOv8 a custom dataset and training is performed. The YOLO-model achieves 99.5% in terms of mean Average Perception (mAP@0.5) for all classes. In summary, as a powerful alternative to work with, YOLOv8 was identified. Standalone or in combination with SIFT a TSR system is shown which can work impacted by several environmental conditions. Based on evaluation of three algorithms an optimized code was developed in which YOLOv8 and SIFT were used in combination as a well performing TSR algorithm, which has above 95% accuracy for each traffic sign tested.

1 INTRODUCTION

As observed in [1], TurtleBot3 can detect different signs using the SIFT algorithm that compares the source image and camera image. Additionally, as a customizable robot [2], TB3 has the flexibility to modify its functionality including Traffic Sign Recognition to improve its accuracy for autonomous driving task. As a starting point the autonomous driving "Autorace-Package" for Robot Operating System is recommended [1]. This work addresses evaluation as well as identification of alternative techniques to SIFT. After extensive research and analysis, one algorithm was found to be superior to all others. It was possible to develop an improved version of the package, which significantly increases accuracy and performance over the previous version. The work is noteworthy because it explores a variety of techniques and concludes with a solution that has a wide range of applications. The only limitation is that we use a ROS version that requires the use of an older version of Ubuntu, but this is neglected in terms of optimal performance of the algorithms. In summary the main objective of this paper is to improve the accuracy of traffic sign recognition and develop an optimized autonomous driving (so called Autorace) package for TB3. This is achieved by implementing various algorithms such as R-CNN, SIFT, ORB, and YOLO, and determining the algorithm that performs best. Once the best algorithm has been found, a technique that best suits our algorithm criteria can be implemented to improve efficiency.

2 METHODS AND ALGORITHMS

For sign recognition, an algorithm is needed that can detect and recognize the traffic sign from the image received by robot. Furthermore, in order to improve the recognition, it is first necessary to understand how it works and what the flow is. As shown in Figure 1, first the TB3 captures the image and then transmits it in raw and compressed form to the system. After that, the system will pass the image to a certain function and this function will detect whether a traffic sign is present in the image or not. If a traffic sign is present, a frame is formed with the coordinates of the image and the modified image is published. So, if a traffic sign is detected, other nodes in the system will know about it.

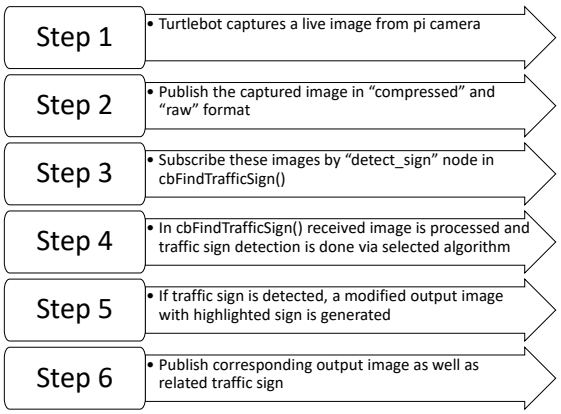


Figure 1: Process of traffic sign recognition.

The focus is on "Step 4", where the input image is processed and traffic sign recognition is performed. This is a usual classification task within machine learning. In the autorace-package [1], the SIFT algorithm is used in combination with the FLANN matcher. The SIFT algorithm extracts key points (features) from the reference images of each traffic sign from local data and also finds key points from the received input image [3, 4]. These key points are then compared with FLANN. The final decision is made to publish whether or not a sign was detected [4].

Figure 2: Traffic signs.

Figure 2 shows which traffic signs shall be recognized using the traffic sign recognition system in TB3. As in Figure 3 an artificial testing environment the Mini-City-Track [1] at Anhalt University of Applied Sciences has all of these signs implemented as part of autonomous driving test.

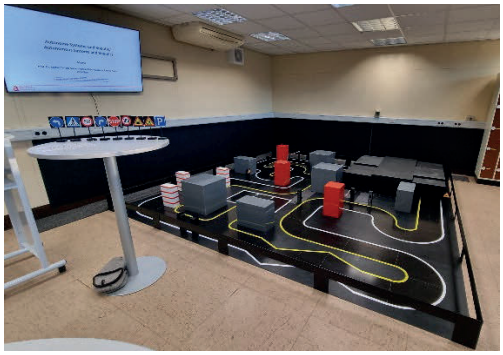


Figure 3: Mini-City-Track at Anhalt University of Applied Sciences.

However, improving efficiency does not mean improving accuracy. In order to increase efficiency, other aspects were also taken into consideration, such as computational time and resources used. As a first impression, SIFT was good in these aspects as it requires the least computational time, but accuracy was limited. It was also observed that accuracy, computation time and resources used were in a devil's triangle relationship with each other, as shown in Figure 4. If one was improved, the other was adversely affected.

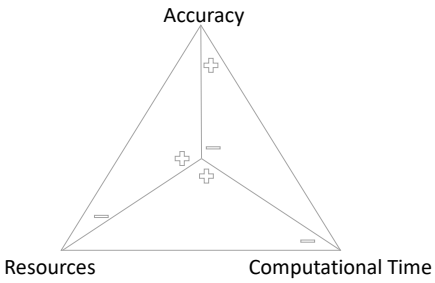


Figure 4: Devil's Triangle.

The ideal state of the Devil's Triangle in Figure 4 is that there is a perfect balance in between with limited resources, manageable computation time and good accuracy, but that is an ideal state. So, usually engineering needs decisions, if accuracy is fully achieved, a lot of resources might be consumed and the computation time probably also increase with complexity, which is not our intention at all. We like to keep it simple like SIFT. This is also important for

the educational purpose, where students have limited time to understand and explore functionality. So, in this paper, we present a TSR system that is close to the ideal state.

2.1 Algorithm Analysis

Various algorithms exist for use in terms of object detection, yet the current cutting-edge technique is "Convolutional Neural Networks (CNN)." [5] CNN serves as a foundation for deep learning, wherein one algorithm handles feature extraction and comparison.

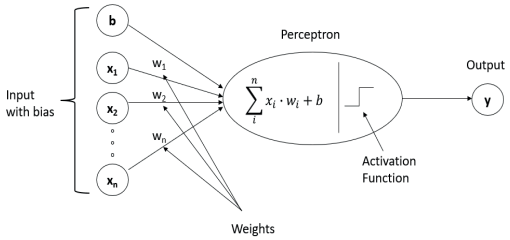


Figure 5: Formula for Neural Networks.

Figure 5 illustrates the basement of Neural Networks – the perceptron is equivalent to a simplified neuron in the human brain. In the perceptron, the inputs are multiplied by their respective weights, and the sum of these value is lead to a (non-linear) output with the assistance of an activation function. Based on this simple representation a network of Neurons consisting of several Neurons in several layers are used where the connection between Neurons can be varied in all directions. Out of that many possible neural network structures can be created, like CNN.

Usually two versions of CNN are used for TSR tasks, which differ based on the number of stages used for detection - One-stage detection, where object classification and bounding boxes are acquired simultaneously, and two-stage detection, where sign area is identified as bounding box within picture first and object classes recognized in a second stage from these. This study compares these deep learning techniques with each other and classical machine learning techniques. The analysis results will determine the most effective algorithm among the currently available techniques for object recognition.

2.1.1 YOLO (You Only Look Once)

According to [6] "YOLOv3: An Incremental Improvement", the YOLO algorithm encodes contextual information about classes and their appearance implicitly during both training and test periods. Hence it can be asserted that YOLO is a single-stage detection system employed by using CNN as the primary principle of detection. YOLO is currently the state-of-the-art technology in terms of having lower background error rates compared to other algorithms. Furthermore, it is easy to comprehend and execute [6]. It also features a range of models to adjust the complexity from "n" to "xl", enhancing accuracy with a superior model [7,8]. As it is a single-stage system, the detection precision is lower, but it is more suitable in real-time systems because it takes less time to detect [8]. However, this algorithm has a drawback - while training a custom dataset, it requires a significant amount of memory and storage space. However, it is currently considered to be the swiftest and most precise algorithm for identifying objects [8].

2.1.2 R-CNN (Region based CNN)

This algorithm operates in two stages, beginning with a proposal of the region of interest where an object may be located. The output of the first stage is then sent to the second stage to classify the object within the designated region of interest. The resulting outcome is a bounding box surrounding the object, along with its classification within that box [9]. It is considered one of the most precise algorithms for detecting objects, as it processes various layers of an image, from the input layer to the hidden layers, all of which are interconnected [10]. The process for training an R-CNN model begins by identifying the region of interest, followed by convolution, non-linearity (using ReLu), and ultimately, maxpooling. This multi-stage approach provides additional features for detection, while the use of maxpooling ensures that it does not consume excessive storage space. Furthermore, a sizeable image can be trained by reducing layers following each convolution and maxpooling. There are also advanced versions of R-CNN available, such as Faster R-CNN, which aim to rectify its shortcomings [10,11]. Despite its various versions, R-CNN is time-consuming to train a model and not suitable for Real-Time systems as it takes a few seconds for detection due to its two-stage detection process [12].

2.1.3 SIFT (Scale-Invariant Feature Transform)

This is a machine learning algorithm used to extract features from a given image. With the help of a feature matcher, the image can then be classified using key points fetched by SIFT. As its name suggests, the algorithm is not affected by the scale or

rotation of the image [3]. It can extract key points even from the smallest image [3]. The traffic signs are automatically annotated by the SIFT algorithm, without the need for human or manual intervention. This is advantageous compared to R-CNN and YOLO [13]. The method is not significantly affected by image size, lighting conditions, or rotation. The main issue with TB3 and SIFT is that it requires an image of the same quality as the reference image.

2.1.4 ORB (Oriented FAST and Rotated BRIEF)

ORB may be employed as an alternative to the SIFT algorithm, resulting in improved efficiency. ORB is constructed using established FAST key point detection and BRIEF descriptors. Because the FAST method is utilized, key points can be detected more rapidly and efficiently with the aid of the BRIEF descriptor. This modified version of the FAST key point decider is used for vision tasks and has superior key point detection abilities compared to the SIFT algorithm [14]. When combined with the BFMatcher [15], it is suitable for traffic sign recognition. Additionally, it is a quick algorithm that requires less computational time than others. However, it operates differently in various environments (background dependency), best performance is reached by using the same environment as the reference image.

2.2 Algorithm Selection

Considering the task and with brief analysis on each algorithm of interest, we decided which algorithm can be suited best for the task. So, YOLO and ORB might be good fit and can be tested/ compared further along with SIFT. Reason for choosing YOLO is the promising accuracy shown in various applications with low computational time [7, 8, 12], and ORB might be a more comparable approach but better version of SIFT as it is said to be faster than SIFT and more accurate also [14]. R-CNN is highly accurate but when working with real-time systems it cannot be used as it is having two stage detection which affects the time used for detection of sign adversely. But along with YOLO and ORB, SIFT will also be implemented as it is the default algorithm for comparison. Furthermore, combinations of these three algorithms are possible and might also be implemented for test, such as YOLO with SIFT or YOLO with ORB.

3 CRITERIA

After carrying out the theoretical analysis, it is necessary to apply the algorithms on TB3 to determine performance and to judge the algorithms. To assess their efficiency, specific criteria must be established:

1) Accuracy for each individual sign;
2) Overall-accuracy for all signs;
3) Calculation/ processing time (Computational performance and Complexity).

Judgement shall be done by TP, TN, FP, FN [16] as with these four categories all the possible outcome can be measured and placed in at least one of these categories, which is helpful to determine the accuracy of the system.

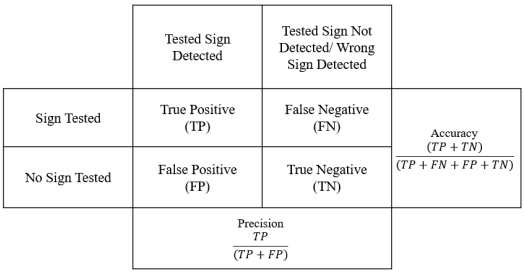


Figure 6: Evaluation Criteria [16].

Figure 6 illustrates. that True Positive is when system detects the sign that was showed, True Negative is that system detects no sign and no sign was shown. False Positive is when system detects a traffic sign even though there is no traffic sign shown, and False Negative occurs when wrong or no traffic sign is detected even though traffic sign is shown. To calculate accuracy and precision two formulas were used from [16] and illustrated in Figure 6. To measure precision, true positive is divided by summation of true positive and false positive, and to calculate accuracy, summation of true positive and true negative is divided by summation of all the four aspects. Furthermore, it is crucial to consider certain impacts when testing the accuracy of each sign individually as well as collectively. Firstly, environmental impacts have to be considered during sign recognition. Light is one of these environmental impacts, for example, the system might be trained in an environment with more or less illumination compared to the environment where the system is utilized. Therefore, testing should be conducted while keeping environmental constraints in mind. Secondly, the positioning of signs is crucial as they are not always in the same place; they can be located

on the right, left, top, bottom, near or far. To overcome this problem, testing the signs in different positions is necessary. Furthermore, it is important to note that for real-time systems, computational time alone is not sufficient. Instead, it is necessary to use it to determine how well the system can detect signs whilst in motion and up to which speed it can detect them without any issues.

4 EXPERIMENTAL SETUPS

The algorithms are tested on the TB3 using multiple test-cases, each with its own significance for fulfilling different criteria.

4.1 On-Table and Mini-City-Track Test

In On-Table test TB3 is placed in laboratory on the table to test functionality at defined conditions. During Mini-City test, the TB3 is placed on a track with predefined tasks and signs.

4.1.1 Environment Independency

To test this, the robot is placed on a table to assess its environment independence, mainly focusing on background interferences, as illustrated in Figure 7. The outcome is recorded, analyzed and a decision made accordingly. The traffic sign recognition system is activated after TB3 was placed on table to evaluate the signs presented in front of it.

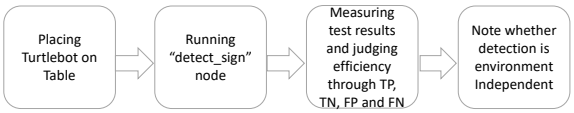


Figure 7: Environment independency test procedure.

4.1.2 Dynamic Positioning Test

There are two dynamic positioning tests: the first involves a moving traffic sign while the TB3 remains stationary, and the second involves the robot in motion on our Mini-City track. For the first test, robot is placed on table with the traffic sign recognition system activated. Different signs are then tested in a continuous manner, going from left to right, up and down, and vice versa. Results are recorded to determine whether the traffic sign recognition system operates independently in positioning. In the second version, TB3 is placed on the track and the traffic sign recognition system is triggered. Additionally, the lane detection function is activated, enabling the robot to move automatically on the signed roads. It shall be determined whether the system can detect signs when the TB3 is in motion or not.

4.1.3 Track Range Test

To determine if the TB3 can effectively detect signs on a track from short or long distances (greater and less than 12cm). This test is used for the evaluation of whether sign detection is affecting the real-time system by means of detecting signs within a required time equivalent to the distance of sign and the related reaction to it from TB3 functionality.

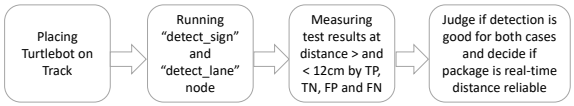


Figure 8: Range test.

Figure 8 outlines the procedure for conducting the test. Prior to the assessment, markings at precisely 12cm in front of all signs are made. These markings serve to determine the accuracy of the robots sign detection prior to and after crossing the mark. The findings from both criteria are combined to make a final determination.

4.1.4 Computational Power Testing

For our case the best method for measuring computational power is to measure the time required to detect a sign.

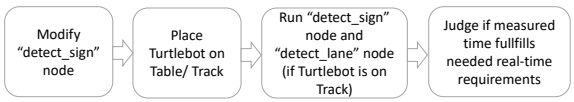


Figure 9: Computational Time Test On-Table/Track.

Figure 9 outlines the method for measuring computational time. Initially, modifications are made to the "detect_sign" node by implementing code for time calculation. The code is inserted after accessing the input sign from the TB3 camera to start the recording process.

```
import time as t
start_time = t.time()
```

Secondly, to measure the time required, following code is added after the node publishes the output image back to the topic it subscribed from.

```
end_time = round(t.time() -
start_time,2) * 1000
print(end_time, "ms")
```

The TB3 is placed on the table or track and the traffic sign recognition system is turned on to detect the sign, by this change the time required for detecting the sign is now also recorded.

5 IMPLEMENTATION AND TESTING

5.1 Implementing Algorithms

The first modification being executed involves altering the sign detection algorithm. As depicted in Figure 10, apart from SIFT, YOLO and ORB algorithms were implemented on the TB3 to assess their effectiveness regarding TSR.

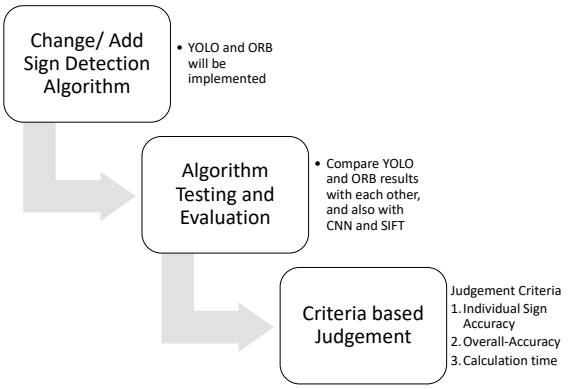


Figure 10: Steps to decide algorithm.

5.1.1 Implementing YOLO for TSR

YOLO is the fastest object detection algorithm [8], although implementation of YOLO is not complex but training custom model on custom dataset is time-consuming [8]. For doing it a properly there are certain steps which are needed to be followed.

5.1.1.1 Training YOLO on a Custom Dataset

The initial stage involves installing the "Ultralytics" library to use in our case YOLOv8. After installation, import the library into the node for application purposes. To train the model, a custom dataset is necessary. To create a dataset, a multitude of images, showing various backgrounds under different lighting conditions, and in different locations are needed, as shown in Figure 11.

Figure 11: Creating images for TSR dataset.

After creating the image dataset, annotation of dataset is necessary. Using online tools, such as Roboflow [17] and CVAT [18], the images can be annotated. Specifically, for this paper, CVAT was utilized. A new project was created, with all necessary classes/labels specified. Afterwards, individual labels are generated for each class and relevant images are assigned to each label. To summarize, the images are manually annotated by dragging a labelled box to the location of the traffic sign within the image.

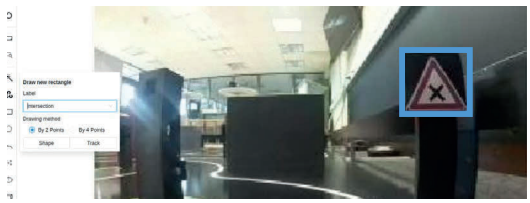


Figure 12: Annotation of sign (image is recorded on Mini-City-Track) with CVAT.

When the images are annotated as displayed in Figure 12, with a box outlining the traffic sign in each image, the data is exported as a dataset, selecting YOLO version to ensure compatibility. Output consists of a label text file corresponding to each image which is located in "labels" directory, while all the images are copied to the "images" directory.

A "config.yaml" file is generated in the same environment. This file specifies the path of the training dataset and defines all the classes/ labels. Within Python [20], which is used for training the YOLOv8n model on our custom dataset, this file is accessed. The model is trained using the "train" function that is pre-defined in the ultralytics library.

```
model = YOLO("yolov8n.pt"
results =
model.train(data="config.yaml",
epochs=65)
```

Finally, a successful training of the model is indicated by a mean Average Perception (mAP@50) near to 1. The "runs" directory, generated during the YOLO training, can then be utilized as a custom model for our custom dataset. This directory contains

weights, training results, and output images from various validation datasets.

5.1.1.2 Using Pre-Trained YOLO Model

Once the model has been trained, it can then be utilized to recognize traffic signs in images obtained from the TB3. To initiate sign recognition, modification is required to the "detect_sign"-node of autorace-package. The pre-trained YOLO model is imported from the "ultralytics" library and subsequently loaded as the model.

```
from ultralytics import YOLO
model_path = os.path.join('.',
'runs', 'detect', 'train',
'weights', 'best.pt',)
model = YOLO(model_path)
```

Afterwards, the YOLO model's "predict()"-function is utilized on the input image to detect signs, and the resulting tensor is analyzed to determine whether any signs were detected. If a traffic sign is detected, the coordinates for bounding box, the classification of the box and its associated confidence are extracted from the tensor. If the confidence exceeds a predetermined threshold, the output image, along with the derived information, is transmitted back to the TB3.

```
results =
model.predict(cv_image_input)
result = results[0]
box = result.boxes[0]
cords = box.xyxy[0].tolist()
sign_ID = box.cls[0].item()
confidence = box.conf[0].item()
xmin,ymin,xmax,ymax =
int(cords[0]), int(cords[1]),
int(cords[2]), int(cords[3])
```

5.1.2 Implementing ORB

ORB is a better form of SIFT algorithm, it almost works in similar pattern as SIFT [14]. It is having reference images for each sign and these images are used to find base key points for each sign and later these key points are matched with the help of BFMatcher to the key points of the input image [15]. It uses less computational time than YOLO as feature extraction of all reference images and initialization of ORB and BFMatcher was done just once before input image is subscribed. Once, one reference image for each sign is taken, “OpenCV2” [19] library is imported in the node because it is used as our base library for machine learning feature extraction. First, all the reference images are loaded in the node and afterwards, key points are fetched from each reference image with the help of ORB.

```
self.orb =
cv2.ORB_create(nfeatures=2000)
self.img2 = cv2.imread(dir_path +
'stop.png',0)
self.kp2, self.des2 =
self.orb.detectAndCompute(self.img
2,None)
```

Once, the key points are fetched from the reference images, the input image subscribed from the TB3 is processed. Therefore, the feature matcher is initiated, in this case “BFMatcher” is used to compare key points of reference images to current image. If a sign is detected a bounding box is drawn on the input image along with which sign it is. Finally, the output image with the bounding box on it is published.

```
self.bf = cv2.BFMatcher(
kp1, des1 =
self.orb.detectAndCompute(cv_image
_input,None)
matches2 =
self.flann.knnMatch(des1,self.des2
,k=2)
```

6 EVALUATION OF PERFORMANCE

Test-cases were formed in a way that all criteria are given for judging efficiency of the system regarding defined task of TSR for autonomous driving on Mini-City track. Within following work each traffic sign (as shown in Figure 2) together with Accuracy and Precision criteria gets a specific abbreviation:

P – Parking;	T – Tunnel;	S – Stop;
I – Intersection;	C – Construction;	50 – Speed 50;
100 – Speed 100;	Li – Traffic Light;	L – Left;
R – Right;	A% – Accuracy;	P% - Precision.

Three test-cases have been selected which focus on both, accuracy and computational time required. All needed changes have been embedded in the autorace-package code. “On-Table test” was conducted on ORB, SIFT, YOLO and YOLO with SIFT. Further, the one with the best results was taken in consideration for “On-Track test” to get to a final decision.

Table 1: Test-Cases.

Test	Description
On-Table: Accuracy	To test accuracy of TB3 through TP, TN, FP, and FN 1) TB3 is placed on a table, so environment/ background independency is tested. 2) TB3 is not moving but the sign shown to robot is moved to different positions like left, right, up, and down.
On-Track: Accuracy	To test accuracy of TB3 through TP, TN, FP, and FN 1) TB3 is placed on the Mini-City track and it performs sign detection along with lane detection, by that motion test is done. 2) Due to moving, signs can be detected from long and short distances. Range is defined as long range for above 12cm and short range for less than 12 cm.
Processing: Time	This test is done during On-Table and On-Track testing, by determining computational time in detect_sign node

6.1 On-Table Results

This test was mainly focused on accuracy in limited environmental conditions as name suggests. This test is conducted on YOLO, ORB and SIFT and the one with the best accuracy will be considered for further testing.

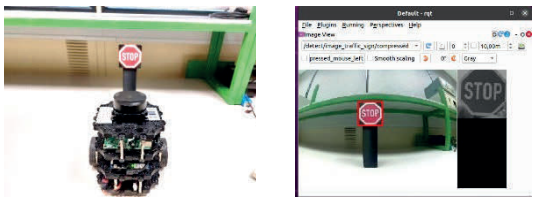


Figure 13: TurtleBot3 during On-Table test.

Figure 13 represents the setup of On-Table test, in which the TB3 is placed on the table in the laboratory and one after the other sign is placed in front of it as shown in the image on the left. Figure 13 right side shows in a ROS rqt-view what TB3 recognizes while On-Table test.

6.1.1 YOLO

Results from Table 2 clearly state that YOLO is having 100% accuracy for five signs out of ten, and 90% or above accuracy for remaining signs except traffic sign “Left” and “Right.”

Table 2: On-Table test YOLO.

	P	T	S	I	C	50	100	Li	L	R
TP	10	10	10	10	9	10	9	9	7	5
FP	0	0	0	0	0	0	0	0	0	0
TN	2	1	1	3	2	2	1	1	3	2
FN	0	0	0	0	1	0	1	1	3	5
A%	100	100	100	100	92	100	91	91	77	58
P%	100	100	100	100	100	10	100	100	100	100

6.1.2 ORB

It is clear from the results shown in Table 3 that ORB is having less than 40% accuracy for all signs except “Left” sign and it cannot be considered as a sign detection algorithm in the final system.

Table 3: On-Table test ORB.

	P	T	S	I	C	50	100	Li	L	R
TP	2	1	3	0	0	2	2	0	8	2
FP	0	0	0	0	0	0	0	0	4	0
TN	2	1	1	2	2	2	1	1	0	2
FN	8	9	7	10	10	8	8	10	2	8
A%	33	18	36	16	16	33	27	9	57	33
P%	100	100	100	100	100	100	100	100	66	100

6.1.3 SIFT

As shown in Figure 14 ORB as well as SIFT uses reference images.

Figure 14: Reference images used for SIFT.

SIFT detects key points in each reference image and further matches those with the current input image’s key points. It is observed in Table 4 that SIFT is having better accuracy than ORB in all signs and in case of “Right” it is having better accuracy than YOLO and for “Left” equivalent to YOLO.

Table 4: On-Table test SIFT.

	P	T	S	I	C	50	100	Li	L	R
TP	1	7	10	8	9	10	8	5	8	10
FP	0	0	0	0	0	0	0	0	2	0
TN	1	1	2	2	2	2	1	5	3	2
FN	9	3	0	2	1	0	2	5	2	0
A%	18	73	100	83	91	100	82	66	73	100
P%	100	100	100	100	100	100	100	100	80	100

However, Figure 15 left side shows also clearly, that partly not all the key points are set to the sign itself, in the shown case one was set to the bottom of the stand.

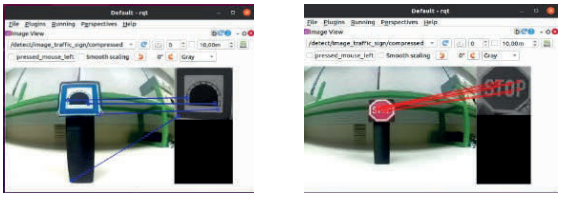


Figure 15: Results of SIFT.

6.2 Optimized Approach

Based on all the gathered experience above and in combination of all techniques suggested, a final way of traffic sign recognition system was designed, which is using YOLO as the primary algorithm for detection of all signs and SIFT as secondary instance which is used only for verification of "Left" and "Right" traffic signs. This way the weakness of YOLO for these two signs shall be vanished. Due to that ORB is not accurate for our use case at all as well as it shows strong environment dependency, it was not further considered. Figure 16 shows the implementation of optimized TSR system. First, the TB3 will start its camera for publishing its video stream. Then, TSR will read frames until and unless system is stopped using "CTRL+C" from terminal. The TSR system starts to process these frames The object detection done by YOLO is detecting trained signs, if traffic sign is detected it checks if identified object is "Left/Right" sign or any other sign, else it will publish no sign detected. For the case of "Left/Right" sign it will cross verify it with the help of SIFT, else it will directly publish the detected sign without using this instance. If SIFT also confirms the same direction like YOLO, the system publishes that sign, else system will publish no sign detected.

6.2.1 Implementing YOLO with SIFT

Quickly, both libraries "Ultralytics" and "OpenCV2", supporting YOLO and SIFT, have been added to the node. The pre-trained model is loaded for YOLO and SIFT is also initiated. Afterwards, for SIFT "Left" and "Right" sign's reference images are loaded and key points are fetched from those two images. FLANN matcher is initiated and the input image is handled by YOLO model before it is optionally forwarded to the SIFT for cross verification. If it is any sign other than direction signs it will be published directly after YOLO detection is finished.

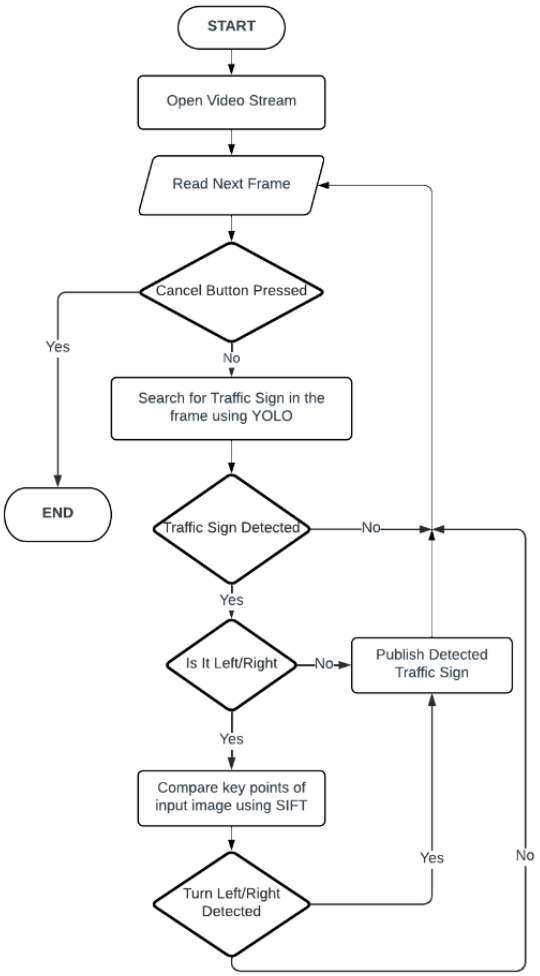


Figure 16: Processing flow of Optimized-TSR based on YOLO combined with SIFT.

6.2.2 On-Table Results YOLO with SIFT

Results from Table 5 clearly state that YOLO with SIFT is having 100% accuracy for five out of ten signs, while for others it is higher than 80%.

Table 5: On-Table test YOLO+SIFT.

	P	T	S	I	C	50	100	Li	L	R
TP	10	10	10	10	9	10	9	9	8	8
FP	0	0	0	0	0	0	0	0	0	0
TN	2	1	1	3	2	2	1	1	3	2
FN	0	0	0	0	1	0	1	1	2	2
A%	100	100	100	100	92	100	91	91	85	83
P%	100	100	100	100	100	100	100	100	100	100

Hence, it can be said that it is best result reached for On-Table test and that is why it is taken for continuing with On-Track test.

6.3 On Track Results

The On-Track test exclusively evaluates the YOLO with SIFT implementation due to that it has demonstrated the best results among all algorithms in the On-Table test. The TB3 is positioned on the Mini-City track during this test to judge dynamic positioning and range dependencies, as well as the computational time when capturing images during motion. The main result of this test is to determine if the implementation is feasible to handle the given real-time task where it has to follow the lane with the aid of a lane detection node while detecting the signs at fixed positions. Obviously, detection must be finished in front of a sign so that other nodes can be called and processed in time. Earlier detection results in more time for processing other tasks during autonomous driving. However, if signs are detected to early, functions might be called that are not yet relevant. Figure 17 illustrates the setup of the On-Track test, where the TB3 moves and upon approaching signs, it detects and notifies the system.

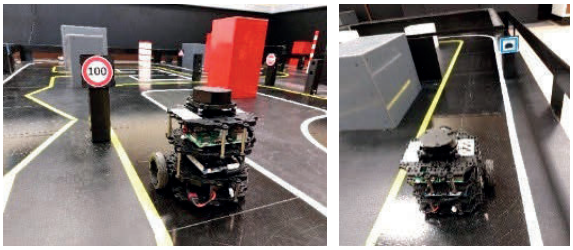


Figure 17: TurtleBot3 during On-Track test.

From Table 6 it is observed that YOLO with SIFT is having 100% accuracy for eight out of ten signs and for remaining two signs it is having above 90% accuracy. If compared to Table 5 “On-Table test YOLO+SIFT” shows that YOLO with SIFT shows better result in artificial-world conditions of Mini-City.

Table 6: On-Track test YOLO+SIFT.

	P	T	S	I	C	50	100	Li	L	R
TP	10	10	10	10	9	10	10	10	9	10
FP	0	0	0	0	0	0	0	0	0	0
TN	2	1	1	3	1	2	1	1	2	3
FN	0	0	0	0	1	0	0	0	1	0
A%	100	100	100	100	91	100	100	100	92	100
P%	100	100	100	100	100	100	100	100	100	100

6.4 Analysis of Algorithm Performance

As explained in Figure 8 “Computation Time Test On-Track/Table” in Chapter 4 “Experimental Setup” within “detect_sign” node the computational time is recorded.

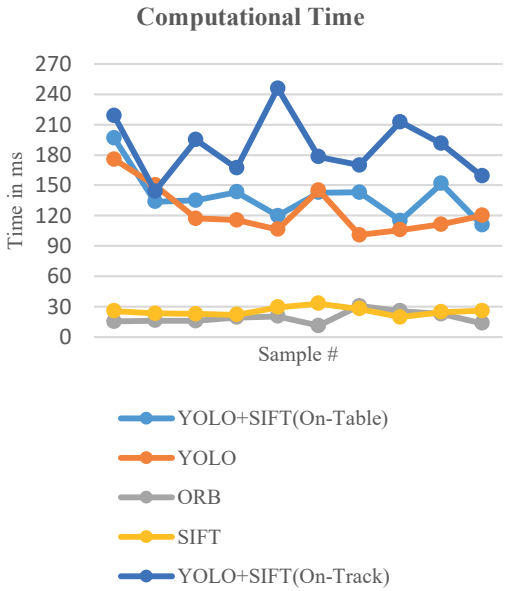
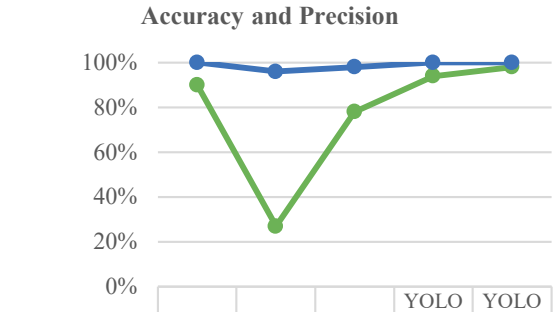


Figure 18: Computational Time Test samples.

At Figure 18, it is clearly visible that YOLO+SIFT is the most time-consuming method for detecting signs and publishing the result. The average time taken for YOLO+SIFT (combining data from both On-Table and On-Track tests) is 163.7ms with a standard deviation of 36.94, highlighting the different effects of using two algorithms in sign detection. YOLO+SIFT takes longer for sign recognition, with an average time of 124.82ms and a standard deviation of 24.20. ORB and SIFT take nearly the same amount of time for traffic sign detection. ORB is a bit quicker with an average time of 19.15ms compared to SIFT's average time of 25.37ms. However, in terms of stability, SIFT is superior with 3.95, whereas for ORB the standard deviation is 5.93. Therefore, it can be concluded that ORB is the fastest algorithm among them, while SIFT yields more consistent outcomes.
In addition, the accuracy and precision of On-Table and On-Track test for different algorithms for all signs was also measured with the help of formula shown in Figure 5 “Evaluation Criteria”.

Average accuracy and precision in Figure 19 demonstrates that the decision to use YOLO for traffic sign detection was justified, as it improved TB3's traffic sign recognition accuracy by 12% to

90% compared to the default SIFT algorithm. Although the ORB algorithm was also implemented, where Figure 17 demonstrates that ORB has the shortest computational time at 19.15ms but with a low performing average accuracy of only 27% and a 4% loss in precision too. Table 2 reveals a lack of accurate detection for traffic signs such as "Left" and "Right" for YOLO.



	YOLO	ORB	SIFT	YOLO +SIFT (On-Table)	YOLO +SIFT (On-Track)
Accuracy	0.90	0.27	0.78	0.94	0.98
Precision	1.00	0.96	0.98	1.00	1.00

Figure 19: Comparison of average accuracy and precision for each algorithm.

The combination of YOLO +SIFT achieved the highest accuracy of 98% on the track, again with perfect precision for traffic sign detection with the effort of a long computational time of 124.82ms. The optimized integration of YOLO with SIFT resulted in 94% accuracy for table and 98% accuracy for track. Whereas most probably the fixed position in comparison to table tests pushed the average 4% up. This approach successfully detected all signs both on and off track with a cost of computational time, which can be managed by reducing frames per second.

7 CONCLUSIONS

The main objective of this paper is to develop an efficient, easy to train and understandable system for traffic sign recognition in educational use. The system should be robust, with minimal errors and operate in real-time. To improve efficiency, several algorithms such as R-CNN, YOLO, ORB, SURF, and SIFT were considered. The advantages and disadvantages of all algorithms, including SIFT, were analyzed through theoretical examination. YOLO, ORB and SIFT were selected for testing on TB3 within an artificial environment, the Mini-City. However, selecting the appropriate algorithm alone does not suffice to boost efficiency; various techniques were also utilized. For instance, a detection frame and a fusion of two algorithms were employed. While these measures enhanced accuracy, they also impacted computational time, necessitating the implementation of numerous test cases.

Two tests - the "On-Table Test," which focused on environment independence, floating traffic signs, and computational power criteria, and the "On-Track Test," which evaluated accuracy based on range and dynamic positioning - were conducted and efficiency was additionally tested via computational time. Test cases have demonstrated the effectiveness of the systems and identified the most suitable technique.

- YOLO exhibited superior accuracy and perfect precision compared to ORB and SIFT in various environments, while the last two were only reliable if tested in the same environmental setting as their reference image. In contrast, YOLO can effectively operate environment independent with any background.
- ORB requires the least computational time, whereas YOLO demands extensive resources to function optimally. This is due to YOLO performing recognition tasks after image access, whereas ORB merely matches key points once the input image is retrieved.
- YOLO itself is not entirely accurate for all signs; therefore, SIFT is introduced in combination with YOLO, but solely on signs that are not detected accurately, rather than all images.

One of the primary issues encountered was the high computational power demand of YOLO, leading to latency in the system and preventing it from being real-time. However, YOLO performed complete detection while intrinsic and extrinsic camera threads were working in the background. Simplest way to cover is to decrease the frames per second within autorace-package for TB3.

```
if self.counter % 10 != 0:
        self.counter += 1
        return
else:
        self.counter = 1
```

After considering the findings and analyzing the computational time of various algorithms and techniques, it was decided that the optimal approach would be to utilize YOLO and SIFT. Of course, this approach might not be needed for a simple TSR exploration within educational use, however, it

demonstrates a usual way of implementing assistant systems where for example in addition to a camera-based object detection digital-map data is used to validate detected objects or situations. As an alternative the YOLO training process could be made more complex to reach better accuracy for Left/ Right sign.

REFERENCES

[1] "ROBOTIS e-Manual," Online]. Available: https://emanual.robotis.com/docs/en/platform/turtlebot3/autonomous_driving/#traffic-sign-detection, [Accessed Sep. 7, 2023].

[2] R. Amsters and P. Slaets, "Turtlebot 3 as a Robotics Education Platform," in Robotics in Education (Advances in Intelligent Systems and Computing), M. Merdan, W. Lepuschitz, G. Koppensteiner, R. Balogh, and D. Obdržálek, Eds., Cham: Springer International Publishing, 2020, pp. 170-181.

[3] B. Zhong and Y. Li. "Image Feature Point Matching Based on Improved SIFT Algorithm," [Accessed Aug. 22, 2023].

[4] V. Vijayan and P. Kp. "FLANN Based Matching with SIFT Descriptors for Drowsy Features Extraction," [Accessed Aug. 22, 2023].

[5] M. A. A. Babiker, M. A. O. Elawad, and A. H. M. Ahmed, "Convolutional Neural Network for a Self-Driving Car in a Virtual Environment," 2019 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Khartoum, Sudan, 2019, pp. 1-6, doi: 10.1109/ICCCEEE46830.2019.9070826.

[6] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," Apr. 2018, [Online]. Available: https://arxiv.org/pdf/1804.02767.

[7] J. Terven and D. Cordova-Esparza, "A Comprehensive Review of YOLO: From YOLOv1 and Beyond," Apr. 2023, [Online]. Available: https://arxiv.org/pdf/2304.00501.

[8] M. Hussain, "YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection," Machines, vol. 11, no. 7, p. 677, 2023, doi: 10.3390/machines11070677.

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, Jun. 2014 - Jun. 2014, pp. 580-587, doi: 10.1109/CVPR.2014.81.

[10] H. Yanagisawa, T. Yamashita, and H. Watanabe, "A study on object detection method from manga images using CNN," in 2018 International Workshop on Advanced Image Technology (IWAIT), 2018, pp. 1-4, doi: 10.1109/IWAIT.2018.8369633.

[11] O. Hmidani and E. M. Ismaili Alaoui, "A comprehensive survey of the R-CNN family for object detection," 2022 5th International Conference on Advanced Communication Technologies and Networking (CommNet), Marrakech, Morocco, 2022, pp. 1-6, doi: 10.1109/CommNet56067.2022.9993862.

[12] J. Du, "Understanding of Object Detection Based on CNN Family and YOLO," J. Phys.: Conf. Ser., vol. 1004, no. 1, p. 12029, 2018, doi: 10.1088/1742-6596/1004/1/012029.

[13] E. Karami, M. Shehata, and A. Smith, "Image Identification Using SIFT Algorithm: Performance Analysis against Different Image Deformations," Oct. 2017, [Online]. Available: https://arxiv.org/pdf/1710.02728.

[14] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. "ORB: An efficient alternative to SIFT or SURF," [Accessed Aug. 22, 2023].

[15] F. K. Noble, "Comparison of OpenCV's feature detectors and feature matchers," in The proceedings of 23rd International Conference on Mechatronics and Machine Vision in Practice: M2VIP 2016 : Nov. 28-30, 2016, Nanjing, Jiangsu, China, Nanjing, China, J. Potgieter, P. Xu, Z.-S. Zhang, X.-S. Wang, H. Yi, and I. C. o. M. a. M. V. i. Practice, Eds., 2016, pp. 1-6, doi: 10.1109/M2VIP.2016.7827292.

[16] Sh. Nimmisha, "Classification of stages of Diabetic Retinopathy using Deep Learning," 2020, doi: 10.13140/RG.2.2.10503.62883.

[17] "Quickstart - Ultralytics YOLOv8 Docs," [Online]. Available: https://docs.ultralytics.com/quickstart/#use-ultralytics-with-cli, [Accessed Sep. 12, 2023].

[18] S. Ola, Th. Bjørsum-Meyer, A. Histace, G. Baatrup, and A. Koulaouzidis, "Annotation Tools in Gastrointestinal Polyp Annotation" Diagnostics 12, no. 10: 2324, 2022, [Online]. Available: https://doi.org/10.3390/ diagnostics12102324

[19] M. Shoeb, M. Akram Ali, M. Shadeel, and M. Abdul Bari, "Self-Driving Car: Using Opencv2 and Machine Learning," The International journal of analytical and experimental modal analysis (IJAEMA), ISSN 0886-9367.

[20] G. Rossum and F.L. Drake, "Python 3 Reference Manual", Scotts Valley, CA: CreateSpace, 2009.

Scrutinised and Compared: HVG Identification Methods in Terms of Common Metrics

Nadiia Kasianchuk[1,2], Yevhenii Kukuruza[3], Vladyslav Ostash[3], Anastasiia Boshtova[4], Dmytro Tsvyk[5] and Matvii Mykhailichenko[6]

[1]*Faculty of Biology, Adam Mickiewicz University, Uniwersytetu Poznańskiego Str. 6, Poznań, Poland*
[2]*Faculty of Pharmacy, Bogomolets National Medical University, Taras Shevchenko Str. 13, Kyiv, Ukraine*
[3]*Faculty of Biotechnology and Biotechnics, Igor Sikorsky Kyiv Polytechnic Institute, Beresteiskyi Avenue 37, Kyiv, Ukraine*
[4]*Educational and Scientific Centre "Institute of Biology and Medicine', Taras Shevchenko National University of Kyiv, Hlushkova Avenue 2, Kyiv, Ukraine*
[5]*Educational and Scientific Institute of International Relations, Taras Shevchenko National University of Kyiv, Illyenka Str. 36, Kyiv, Ukraine*
[6]*Faculty of Biotechnology, University of Wroclaw, Fryderyka Joliot-Curie 14a, Wroclaw, Poland*
nadkas2@st.amu.edu.pl, evgenku1508@gmail.com, ostash.vladyslav@lll.kpi.ua, a.boshtova@gmail.com, tsvykdima@gmail.com, 341450@uwr.edu.pl

Keywords: Highly Variable Genes, Single-Cell RNA-Sequencing, Differential Expression, Heterogeneity Analysis, Cellular Heterogeneity, Cellular Diversity.

Abstract: Highly variable gene (HVG) identification plays a critical role in unravelling gene expression patterns and understanding cellular heterogeneity in single-cell RNA-sequencing (scRNA-seq) data. A plethora of software packages have been developed for this purpose; however, their comparative performance is yet to be explored. This study addresses this gap by independently evaluating 22 methods from 9 different packages to provide a comprehensive assessment of the HVG identification methods. For such purpose it was deemed necessary to employ a set of common metrics, namely overlap with highly and lowly expressed genes, runtime, and clustering indices (e.g., Calinski-Harabasz, Davies-Bouldin, and ROGUE). The results reveal substantial disparities not only between different methods but also in the performance of a single method across diverse datasets. That is to say, the dimensionality of the provided data, spike-ins, and background noise are some of the key factors influencing the results. These variations underscore the significant impact of dataset characteristics on analysis outcomes. Therefore, consistent consideration of data nature is imperative. The study emphasises the urgent need for a standardised, data-driven assessment framework to ensure reliable and effective scRNA-seq analyses. This work serves as a valuable resource for both scRNA-seq software developers and experimental researchers seeking optimal methods for their investigations.

1 INTRODUCTION

In the era of data-driven medical research, the ways in which biological systems are screened and evaluated have been redefined. Thus, the challenges implied by complexity, heterogeneity, and multidimensionality of data necessitate innovative computational approaches successfully to collect, preprocess [1] and analyse [2, 3, 4, 5] the records.

Therefore, multiomics techniques have been developed as a pivotal tool in addressing contemporary biomedical tasks and challenges. Among the groundbreaking technologies in this domain, single-cell RNA sequencing (scRNA-seq) is of particular standing, as it characterises gene expression patterns at a single-cell resolution [6]. The identification of highly variable genes (HVGs) is crucial within the abovementioned process for profiling cell subtypes, performing dimensionality reduction techniques and unveiling cellular heterogeneity.

A number of methods have recently been put forward to deal with such identification in the scRNA-seq, with many of them being able to extract valuable insights from the data despite the challenges posed. Concurrently, only a minimal number of independent evaluations have been conducted to underline the most optimal methods of identifying

HVGs [7]. Furthermore, there is a lack of documentation regarding which methods are the most suitable for specific types and dimensions of data. Such thorough assessment is vital, given the methods perform ambivalently across diverse datasets, both real-world and synthetically generated [8, 9, 10].

Hence, the purpose of the present study is to establish an efficient and dependable evaluation of existing tools for HVG identification. The groundwork herein provided consists of 5 diverse scRNA-seq datasets being assessed by commonly used methods to distinguish those with the best performance. The study serves as the foundational element in crafting a comprehensive algorithm for the real-time assessment of the methods utilised in the identification of HVGs on the specific dataset required for the exact research. Thence, the researchers will be able to make well-informed decisions on the methods most suitable for their specific objectives, improving accuracy and reliability in scRNA-seq data analysis.

2 MATERIALS AND METHODS

2.1 Data Acquisition and Preprocessing

For the purposes of the present study, publicly available scRNA-seq datasets were collected in such a way that they differ in terms of dimensionality, cell and gene types, ensuring a comprehensive evaluation.

Prior to the analysis, data preprocessing pipeline was implemented to provide quality and comparability of the scRNA-seq data across all the assessed tools. Consistent preprocessing steps were followed for all datasets.

Cells and genes devoid of information regarding expression patterns were excluded from subsequent analyses. Outlier batches were removed only where they significantly impacted detection thresholds. In the datasets with droplet-based sequencing, multiplet droplets and cells with low unique molecular identifier (UMI) counts were omitted (UMI count < 100, false discovery rate threshold 0.001), except when UMI counts had no impact on barcode ranks. Log-normalised counts matrix, size-normalised counts and raw counts normalisation methods were utilised based on the data characteristics within each dataset. Outliers were identified by expression, spike-ins, and mitochondrial gene percentage.

2.2 Selection of HVG Identification Tools

22 methods computed using 9 common packages were chosen for the evaluation. The choice was based on their widespread adoption in the scRNA-seq community, ability to handle large-scale datasets, and capacity to accommodate different data distributions and experimental conditions. Furthermore, the obtained set consisted of both well-studied and not yet assessed packages, so that the comparison could be possible. Most chosen packages were R-based, considering the popularity of this language within biological data science, however Python-written Scanpy was included to cater to the preferences of different researchers and provide a comprehensive evaluation. The methods can be categorised into two groups:

- Variance-based: This group employs metrics related to variance. It selects genes with higher variability under diverse adjustments, assuming their relevance to the dataset's structure. Included methods are M3Drop_Brennecke, Seurat_vst, Seurat_disp, scVEGs, SIEVE_Seurat_vst, SIEVE_Seurat_disp, scLVM_counts, scLVM_log, scLVM_logvar, M3Drop_Brennecke_ERCCs, scLVM_logvar_ERCCs, scLVM_counts_ERCCs, (R-based) and scanpy_seurat, scanpy_cell_ranger (Python-based).
- Distribution-based relies on the dropout rate (prevalence of zeros) or assumption that count data follows a certain distribution. Methods in this category include M3Drop, M3Drop_Basic, ROGUE, ROGUE_n, Seurat_sct, SCHS, scmap, SIEVE_ROGUE, SIEVE_M3Drop (R-based) and scanpy_Pearson (Python-based).

2.3 Assessment Procedure

The process of evaluation involved a systematic test of each tool on the chosen scRNA-seq datasets. To ensure unbiased comparisons, the tools were applied with default parameter settings, and each tool was run on the same computing infrastructure.

2.3.1 Variance-Mean Dependence

Heteroscedasticity, a common issue in single-cell RNA-seq data, can introduce unwanted variability in

the analysis. Therefore, to measure the potential influence of expression mean on the identification of HVGs, we quantified the ratio of HVGs that coincided with the top highly and lowly expressing genes, as well as showed Pearson's correlation between the mean expression values and variance.

2.3.2 Clustering Validation Metrics

It is crucial to ensure that selected HVGs follow some distinct expression patterns and are associated with specific clusters of genes. Therefore, several widely used clustering validation metrics were utilised for assessing the performance of the HVG identification methods.

The Calinski-Harabasz (CH) index assesses the compactness and separation between clusters, yielding higher values for better-defined ones. It is calculated by taking the ratio of the between-cluster dispersion to the within-cluster dispersion, where dispersion refers to the sum of squared distances between data points and their cluster centroids. Higher CH index values indicate better-defined and more compact clusters.

The Adjusted Rand Index (ARI) compares the similarity between different partitions of data obtained from the same clustering method, accounting for chance agreements; higher ARI values indicate better internal consistency and separation of clusters.

The Davies-Bouldin (DB) index measures the average similarity between each cluster and its most similar cluster, aiming for a lower value to indicate better-defined parameters.

The ROGUE metric [11] is tailored for scRNA-seq data and utilises an entropy-based method to quantify the purity of cell clusters. Average Silhouette Width quantifies cohesion and separation of data points within clusters, with higher values indicating well-clustered data.

Additionally, we incorporated the Purity of t-SNE k-means Clustering, that is a useful technique utilised for assessing the quality of generated clusters. For such purpose, t-distributed stochastic neighbour embedding with following k-means clustering was performed on the obtained data. Either pre-existing labels or external references were used for assigning the cell subtypes and the cluster subtype was deemed as the most common cell subtype in the cluster. Purity was calculated as the ratio of cells assigned to correct cluster, the calculation was repeated 3 times and the average was taken into the further analyses.

3 RESULTS AND DISCUSSION

The assessment was performed on 5 publicly available scRNA-seq datasets, carefully chosen to represent different types of biological data. Human datasets, such as Mair [12] (peripheral blood mononuclear cells) and Campbell [13] (brain), and mouse datasets, including Richard [14] (T cells) and Buettner [15] (embryonic stem cells), were obtained using the scRNA-seq R package [16] . The HIV [17] dataset (human HIV and CMV-specific CD8+ T-cells) was sourced from the Single Cell Portal database [18]. Data was thoroughly pre-processed, and all missing values and outliers were excluded from the analyses (Table 1).

Table 1: Dimensionality of datasets before and after the Quality Control (QC).

Name of dataset	Genes (Before QC)	Genes (After QC)	Cells (Before QC)	Cells (After QC)
Buettner	38293	11318	288	87
Mair	499	471	29033	6540
HIV	12122	12122	1559	1073
Richard	46603	28402	46603	528
Campbell	26774	26383	21086	17172

3.1 Heteroscedasticity and Variance-Mean Dependence

As proven by a number of papers, a strong correlation (e.g., 0.951, $p < 0.001$ in Mair dataset) exists between mean expression values and variance (Figure 1) causing the heteroscedasticity. Therefore, variance cannot be used as a direct indicator of HVGs and mean–variance relationships should be given attention during analyses.

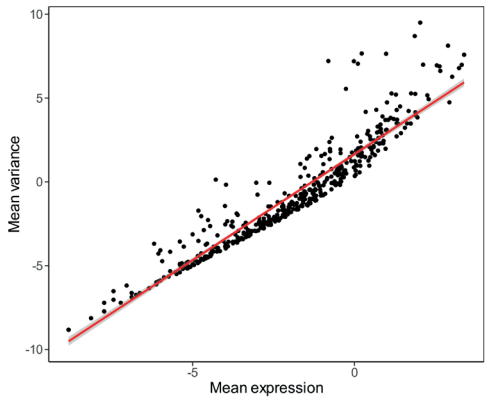


Figure 1: Correlation between the mean expression and variance in the scRNA-seq data assessed on the Mair dataset.

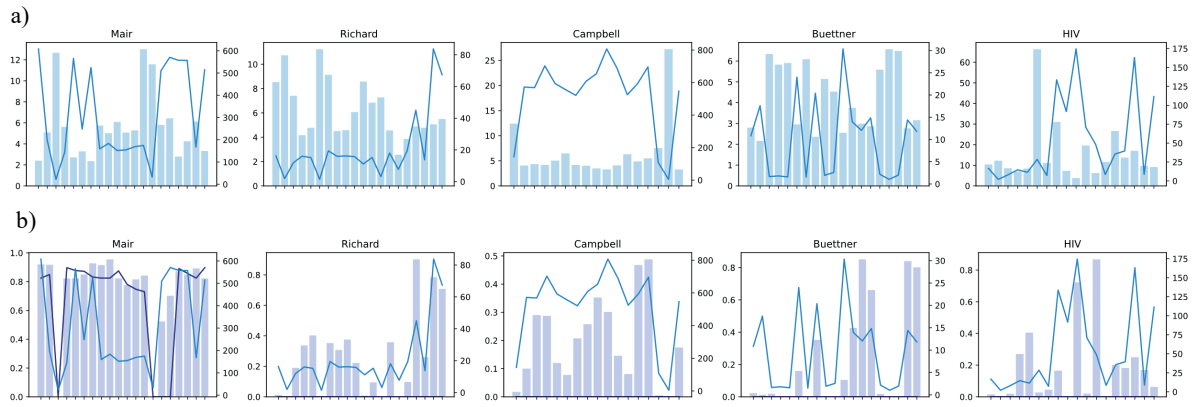


Figure 2: Evaluation of the HVG identification methods based on the quality of clustering and dependence on the heteroscedacity. a) The x-axis displays various tools used for HVG identification. A blue line plot, aligned with the right y-axis, shows the CH index. The histogram, corresponding to the left y-axis, presents the DB results, revealing a general inverse relationship between the CH index and DB values, highlighting differences in clustering efficacy. b) The x-axis again lists the tools. A blue line plot and right y-axis depicts the CH index. A purple line plot, alongside a histogram and left y-axis, shows the percentage overlap of HVGs with HEGs and LEGs. Notably, all datasets exhibited no overlap between LEGs and HVGs, with the exception being the Mair dataset.

Though multiple approaches are implemented to overcome heteroscedasticity, identification of lowly expressed genes (LEGs) still remains a daunting task in most of the methods [19, 20]. That is to say, no overlap was witnessed between lowly expressed genes (bottom x genes sorted by expression, where x is the number of HVGs returned by the respective method) and HVGs in all datasets, except in Mair (Figure 2b). However, such outcome might have been caused by low dimensionality of the dataset in question [17], [21].

The extent of overlap differed significantly across the datasets, underscoring how the methods' effectiveness is greatly influenced by the specific nature and dimensionality of the data. For instance, SCHS showcased substantial highly expressed genes (HEGs) overlap within the Mair and HIV datasets, aligning with the earlier findings [9]. However, within datasets containing a larger number of genes, the observed overlap leaned towards the moderate.

Only scLVM-based methods showed relatively consistent results, usually emerging as leaders in respect of HVG-HEGs overlap. On the other hand, M3Drop-based approaches exhibited relatively low overlap in most cases, possibly suggesting a reduced reliance on the variance (Figure 2b).

3.2 Runtime

In the realm of computational tool evaluation, runtime encapsulates efficiency and performance, shedding light on the methods' scalability, resource utilisation, and overall responsiveness [22].

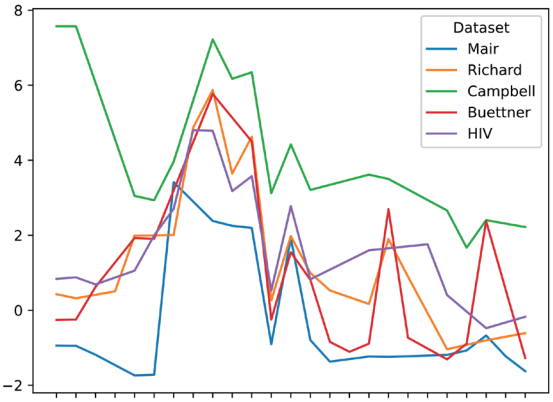
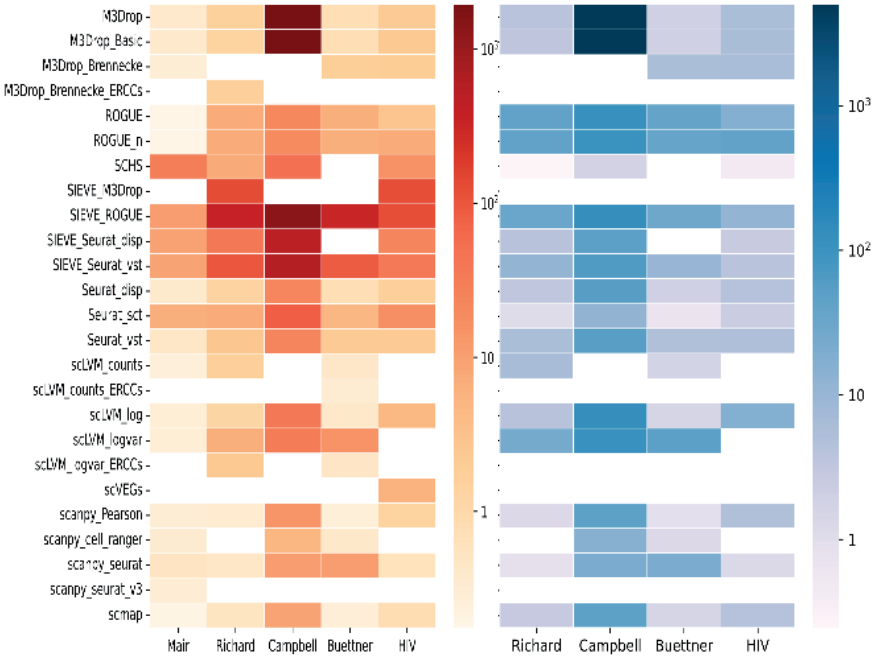


Figure 3: The runtime of the methods as indicated on the x-axis, with the y-axis representing the log-normalized runtime values in seconds.

Overall, Scmap demonstrated the best results in terms of runtime, being within the top three fastest methods across all datasets (Figure 3). Notably, Scmap adeptly handles high-dimensional datasets, exhibiting a 46,6-fold increase in runtime in the Campbell dataset compared to Mair (Figure 4). Scanpy-based methods also exhibited relatively swift execution time among all datasets not relying much on the dimensionality of the provided datasets. Seurat- and scLVM-based approaches also performed with commendable speed. Moreover, the use of External RNA Controls Consortium (ERCCs) significantly enhanced the runtime of the latter (Figure 4).



a) b)

Figure 4: Runtime assessment of the investigated methods: a) Log-normalized analysis time for individual datasets. b) The heatmap visualises the log-normalised augmentation in runtime relative to the Mair dataset, which serves as the baseline with a runtime value of 1.

Intriguingly, SCHS showed satisfactory outcomes across all datasets, other than the smallest Mair, where it performed with a strikingly low speed. M3Drop, on the other hand, distinctively underperformed on the high dimensional datasets. Incorporation of SIEVE substantially slowed down the performance of all methods used for comparison (Figure 3).

3.3 General Clustering Evaluation

Accurately identified HVGs should display distinct expression profiles across various cell populations or conditions. This specificity allows them to serve as valuable signatures for subsequent analyses. Consequently, metrics like cluster purity, heterogeneity and separation play a pivotal role in comparing and assessing different tools in this context.

Considering the high dimensionality and complexity of data discussed, several clustering indices were assessed, assuring non-biased evaluation of the HVG identification methods. Remarkably, no single method exhibited consistent superiority or inferiority across all datasets, nor did any method demonstrate such performance across all indices (Figure 2a).

M3Drop showcased the best results in the low-dimensional Mair dataset, ranking top-3 in CH, DB and ARI metrics, though it struggled while dealing with Campbell data. M3Drop_Brennecke, on the other hand, was among the worst methods on the same Mair dataset. The evaluation of Scanpy_cell_ranger also differed tremendously, as it ended up having one of the lowest scores in Campbell and Buettner, while showing good outcome in Mair. Moreover, ERCCs utilisation heightened both the CH and DB scores in all methods, where it was applied. This intriguing observation warrants deeper investigations to discern the underlying reasons for this phenomenon, as, traditionally, a high CH score signifies improved clustering outcomes, whereas a high DB score indicates the opposite.

Such divergence and somewhat chaos in the scores prove that methods' performance strongly depend on the data used for analyses [10], [23] and, therefore, some standardised method of choosing the optimal tool should be developed.

3.4 Average Score for Clustering Evaluation

As explained in the previous chapter, systematic evaluation of HVG-based clustering is necessary. However, this task is challenging, since performance

significantly depends on type, volume, and complexity of the data provided. Consequently, an average score was calculated using the z-normalised results of all clustering-based metrics. Furthermore, in all indices except DB, a higher value indicates better clustering characteristics. To account for this, DB values were inverted.

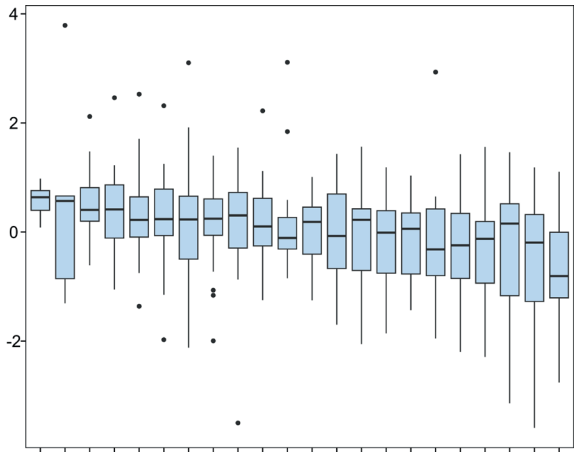


Figure 5: Average score of clustering-based metrics. The x-axis illustrates the various tools, while the y-axis represents the z-normalized average scores of clustering-based metrics.

Overall, the SCHS, scVEGs, and scanpy_seurat_v3 methods exhibited the highest performance (Figure 5) with their scores notably surpassing those of other tools (0,57-0,60 compared to 0,30 in the method ranked 4). ROGUE_n (log-normalised counts), along with the SIEVE-incorporated ROGUE, demonstrated commendable performance, securing ranks 4 and 5. However, the same cannot be said for pure ROGUE, as it achieved an average score of -0,18.

Noteworthy, the incorporation of SIEVE improved performance only of the abovementioned ROGUE, as Seurat_vst and Seurat_disp exhibited a fall off in the average score. Scmap ended up with a moderate result of 0,11 and M3Drop underperformed (averaged score < 0) in all cases.

However, it should be noted, that although the averaged normalised score is a convenient way for the basic comparison of methods, it cannot provide insights into the strengths and weaknesses of each exact method. Therefore, next steps are required to enable systematic overview of the HVG identification tools.

4 CONCLUSIONS

HVG identification is a vital component of single cell RNA-sequencing providing a comprehensive overview of gene expression patterns and a deeper comprehension of cellular heterogeneity. While numerous packages have been developed for this purpose, information pertaining to their performance remains limited: many evaluations have been conducted by the developers of these methods, introducing a potential conflict of interests. Hence, an independent assessment may come in handy for both scRNA-seq software developers and wet researchers who look for a suitable method to employ in their research.

Remarkably, the outcomes exhibited significant discrepancies, not merely among methods themselves, but also in performance of a given method when applied to various datasets. However, these variations are in line with findings from other independent research studies in the field [6]. To name a few, M3Drop exhibited moderate to good performance in small datasets, but with the increase of data dimensionality it strongly underperformed both in terms of clustering and runtime. Similar trends were observed with Scanpy_cell_ranger. Conversely, SCHS runtime increased when analysing the Mair dataset, which has the lowest dimensionality. These divergences underscore the substantial influence of the analysed data's nature on the obtained results. Hence, consistently considering this aspect is crucial, and the development of a standardised data-driven assessment system is imperative to ensure effective and reliable analyses in the field.

ACKNOWLEDGMENTS

With heartfelt appreciation, we extend our gratitude to the Bioinformatics for Ukraine initiative, and particularly to Prof. Serghei Mangul, for inspiring us to continue this research project following the East European Bioinformatics and Computational Genomics School. We also express our deep gratitude to everyone who courageously fight for the freedom of Ukraine and the whole civilised world.

REFERENCES

[1] S. Fedushko, M. Gregus, and T. Ustyianovych, 'Medical card data imputation and patient psychological and behavioral profile construction', Procedia Comput Sci, vol. 160, pp. 354-361, 2019, doi: 10.1016/j.procs.2019.11.080.

[2] M. Marczyk et al., 'Treatment Efficacy Score-continuous residual cancer burden-based metric to compare neoadjuvant chemotherapy efficacy between randomized trial arms in breast cancer trials', Annals of Oncology, vol. 33, no. 8, pp. 814-823, Aug. 2022, doi: 10.1016/j.annonc.2022.04.072.

[3] P. Rzymski, N. Kasianchuk, D. Sikora, and B. Poniedziałek, 'COVID-19 vaccinations and rates of infections, hospitalizations, ICU admissions, and deaths in Europe during SARS-CoV-2 Omicron wave in the first quarter of 2022', J Med Virol, vol. 95, no. 1, Jan. 2023, doi: 10.1002/jmv.28131.

[4] N. Kasianchuk, D. Tsvyk, E. Siemens, and H. Falfushynska, 'Random Forest Algorithm in Unravelling Biomarkers of Breast Cancer Progression'. Proceedings of the International Conference on Applied Innovations in IT (ICAIIT), vol. 11, no. 1, pp. 133-141, Mar. 2023, doi: 10.25673/101930.

[5] N. Kasianchuk, D. Tsvyk, E. Siemens, V. Ostash, and H. Falfushynska, "Genomic data machined: The random forest algorithm for discovering breast cancer biomarkers," in Information and Communication Technologies and Sustainable Development, in Lecture notes in networks and systems. Cham: Springer Nature Switzerland, 2023, pp. 428-443. doi: 10.1007/978-3-031-46880-3_25.

[6] D. Deshpande et al., 'RNA-seq data science: From raw data to effective interpretation', Front Genet, vol. 14, Mar. 2023, doi: 10.3389/fgene.2023.997383.

[7] A. Sonrel et al., 'Meta-analysis of (single-cell method) benchmarks reveals the need for extensibility and interoperability', Genome Biol, vol. 24, no. 1, p. 119, May 2023, doi: 10.1186/s13059-023-02962-5.

[8] S. H. Yip, P. C. Sham, and J. Wang, 'Evaluation of tools for highly variable gene discovery from single-cell RNA-seq data', Brief Bioinform, vol. 20, no. 4, pp. 1583-1589, Jul. 2019, doi: 10.1093/bib/bby011.

[9] Y. Zhang, X. Xie, P. Wu, and P. Zhu, 'SIEVE: identifying robust single cell variable genes for single-cell RNA sequencing data', Blood Science, vol. 3, no. 2, pp. 35-39, Apr. 2021, doi: 10.1097/BS9.0000000000000072.

[10] T. S. Andrews and M. Hemberg, 'M3Drop: dropout-based feature selection for scRNASeq', Bioinformatics, vol. 35, no. 16, pp. 2865–2867, Aug. 2019, doi: 10.1093/bioinformatics/bty1044.

[11] B. Liu, C. Li, Z. Li, D. Wang, X. Ren, and Z. Zhang, 'An entropy-based metric for assessing the purity of single cell populations', Nat Commun, vol. 11, no. 1, p. 3155, Jun. 2020, doi: 10.1038/s41467-020-16904-3.

[12] F. Mair et al., 'A Targeted Multi-omic Analysis Approach Measures Protein Expression and Low-Abundance Transcripts on the Single-Cell Level', Cell Rep, vol. 31, no. 1, p. 107499, Apr. 2020, doi: 10.1016/j.celrep.2020.03.063.

[13] J. N. Campbell et al., 'A molecular census of arcuate hypothalamus and median eminence cell types', Nat Neurosci, vol. 20, no. 3, pp. 484-496, Mar. 2017, doi: 10.1038/nn.4495.

[14] A. C. Richard, A. T. L. Lun, W. W. Y. Lau, B. Göttgens, J. C. Marioni, and G. M. Griffiths, 'T cell cytolytic capacity is independent of initial stimulation strength', Nat Immunol, vol. 19, no. 8, pp. 849-858, Aug. 2018, doi: 10.1038/s41590-018-0160-9.

[15] F. Buettner et al., 'Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells', Nat Biotechnol, vol. 33, no. 2, pp. 155-160, Feb. 2015, doi: 10.1038/nbt.3102.

[16] D. Risso et al., 'scRNAseq: Collection of Public Single-Cell RNA-Seq Datasets', R package version 2.14.0, 2023, doi: 10.18129/B9.bioc.scRNAseq.

[17] R. Fardoos et al., 'HIV specific CD8+ TRM-like cells in tonsils express exhaustive signatures in the absence of natural HIV control', Front Immunol, vol. 13, Oct. 2022, doi: 10.3389/fimmu.2022.912038.

[18] 'Single Cell Portal'. Accessed: Jul. 22, 2023. [Online]. Available: https://singlecell.broadinstitute.org/single_cell.

[19] H.-I. H. Chen, Y. Jin, Y. Huang, and Y. Chen, 'Detection of high variability in gene expression from single-cell RNA-seq profiling', BMC Genomics, vol. 17, no. S7, p. 508, Aug. 2016, doi: 10.1186/s12864-016-2897-6.

[20] C. A. Vallejos, J. C. Marioni, and S. Richardson, 'BASiCS: Bayesian Analysis of Single-Cell Sequencing Data', PLoS Comput Biol, vol. 11, no. 6, p. e1004333, Jun. 2015, doi: 10.1371/journal.pcbi.1004333.

[21] F. A. Wolf, P. Angerer, and F. J. Theis, 'SCANPY: large-scale single-cell gene expression data analysis', Genome Biol, vol. 19, no. 1, p. 15, Dec. 2018, doi: 10.1186/s13059-017-1382-0.

[22] A. Tyryshkina, N. Coraor, and A. Nekrutenko, 'Predicting runtimes of bioinformatics tools based on historical data: five years of Galaxy usage', Bioinformatics, vol. 35, no. 18, pp. 3453-3460, Sep. 2019, doi: 10.1093/bioinformatics/btz054.

[23] P. Brennecke et al., 'Accounting for technical noise in single-cell RNA-seq experiments', Nat Methods, vol. 10, no. 11, pp. 1093-1095, Nov. 2013, doi: 10.1038/nmeth.2645.

APPENDIX

For access to the code, supplementary data, and results pertaining to this research, please visit our GitHub repository at https://github.com/RI3NO/HiVaGe.

Demographic Bias in Medical Datasets for Clinical AI

Bojana Velichkovska, Sandra Petrushevska, Bisera Runcheva and Marija Kalendar
Faculty of Electrical Engineering and Infomation Technologies, "Ss. Cyril and Methodius University" in Skopje, Rugjer Boshkovikj 18, Skopje, North Macedonia
{bojanav, kti1772019, kti1842019, marijaka}@feit.ukim.edu.mk

Keywords: Artificial Intelligence, Machine Learning, Gender Bias, Age Bias, Demographic Bias, Medical Datasets.

Abstract: Numerous studies have detailed instances of demographic bias in medical data and artificial intelligence (AI) systems used in medical setting. Moreover, these studies have also shown how these biases can significantly impact the access to and quality of care, as well as quality of life for patients belonging in certain under-represented groups. These groups are then being marginalised because of stigma based on demographic information such as race, gender, age, ability, and so on. Since the performance of AI models is highly dependent on the quality of data used to train the algorithms, it is a necessary precaution to analyse any potential bias inadvertently existent in the data, in order to mitigate the consequences of using biased data in creating medical AI systems. For that reason, we propose a machine learning (ML) analysis which receives patient biosignals as input information and analyses them for two types of demographic bias, namely gender and age bias. The analysis is performed using several ML algorithms (Logistic Regression, Decision Trees, Random Forest, and XGBoost). The trained models are evaluated with a holdout technique and by observing the confusion matrixes and the classification reports. The results show that the models are capable of detecting bias in data. This makes the proposed approach one way to identify bias in data, especially throughout the process of building AI-based medical systems. Consequently, the proposed pipeline can be used as a mitigation technique for bias analysis in data.

1 INTRODUCTION

There exit numerous factors which contribute to or exacerbate disparities in healthcare, as are implicit and explicit biases which imbibe discriminatory practices based on demographic information as race, ethnicity, gender, or age [1]. With biased practices preserving, patients can receive subpar care quality, which can range from delays in admission and poor treatment to inaccurate diagnosis and potential for worsened health conditions [2].

The impact of these issues largely affects underrepresented groups, and these (un)intended consequences even impede academic performance as medical professionals find themselves unable to treat certain populations. In example, dermatologists have spoken of their inability to accurately diagnose diseases in patients of colour due to under-representation of certain populations in medical textbooks [3]. Consequently, five-year melanoma survival estimations show the survival rate for Black patients is only 70% compared to the 94% for White patients [4]. Compared to the self-awareness of dermatologists, there is a different side to medical personnel, as shown by [5], where it is illustrated that physicians are significantly less likely to recommend bypass surgery for Black compared to White patients. The contributing factor in these decisions was physicians believing Black patients to be less educated and, therefore, less likely to adhere to necessary activity post-surgery.

Moreover, personnel biases extend to disability attitudes [6], with 83.6% of healthcare providers having a preference for able-bodied patients. Socioeconomic status is another aspect in which medicine is biased, and patients of lower status are likely to have worse self-reported health and at a risk of multimorbidity [7], in addition to having limited access to health care and being at a greater risk for substandard care [8].

The biases are not limited to preferences only, and extend to assumptions based on demographic information which personnel use when treating patients. The authors of [9] identify gender bias in patient-provider encounters and treatment decisions, with dichotomous depictions of "brave men" and "emotional women". The study also found that

physicians are likely to attribute woman's pain as a product of a mental health condition rather than as a physical condition. Medical personnel disregarding patients' conditions can lead at the very least to delays in diagnosis. One example is [10], which found that women wait longer on average for a diagnosis compared to men in 72% of cases. Worst case scenarios can result in increased risk of death, e.g., how lack of awareness of the impact of heart attacks on women contributes to higher rates of females dying from heart attacks [11]. Healthcare professionals are less likely to recommend older patients for invasive or aggressive procedures denoting the choice as a "compassionate" approach even if said decision impacts life quality and expectancy in these patients [12].

Despite efforts to address and mitigate biased practices, health inequities persist, and infinitely worse get propagated in medical datasets and AI models which impact large populations. An algorithm was found to be racially biased since it used medical costs as a proxy for care needed, and consequently assigned the same level of risk to Black and White patients, even though the Black patients were in a worse medical condition [13]. A study of an algorithm for abnormalities in chest X-rays showed that highest rate of underdiagnosis exists in young females [14]. Another algorithm, which aimed to help with in-home care for patients, was found to recommend extreme cuts in cases of disabled patients, resulting in reduced quality of life and increased hospitalisation [15].

As integration of AI in medical systems is expected to increase in the upcoming years, it is necessary to address and resolve biased issues in order to limit negative impact, as well as understand where the bias originates in order to reduce the chances of propagating said bias into production stages, and thus, mitigation strategies will be necessary. Previous examples demonstrate that one potential source of bias for AI models can be the data used for the research and its distributions, as shown in [16] where the authors show the impact of gender imbalance in medical imaging datasets in computer-aided diagnostic tools. Additionally, the data used is the driving force for the algorithms, as they extrapolate information from said data in order to understand the problem and arrive at a decision.

Since the basic foundation for AI systems is the data, we wanted to investigate whether data bias is visible and easily discernable by the algorithms even when confounding variables are excluded from the training data. That is to say, we investigate whether potential biased issues can be detected with simple analysis of the data itself. However, as all data can be a subject to bias, medical datasets are not excluded from the influences of biased medical personnel or biased decisions in real practice. Moreover, even though it is necessary for developers to thoroughly investigate trained models before their active use, in many cases hidden (or implicit) biases are not observed before models are deployed. This results in biased real-world applications, which impact large populations [13, 14]. Normally, biases arise from using confounding variables, however bias can be present even when confounding variables are excluded from research.

For that reason, we wanted to investigate whether implicit biases can be found in data points where they should not exist, namely, measurements from bedside monitors. Therefore, we analyse bias from two demographic aspects, age and gender, using machine learning (ML) algorithms. The model is derived on 80% of the data, whilst the performance is evaluated from a holdout of 20% using a classification report [17] and confusion matrix as metrics [18].

Previous papers have shown both gender [19, 20, 21] and age [22] differences in biosignals. Moreover, ML algorithms have been used to predict age and gender from iris biometrics [23, 24]. ML has also been used for racial bias analysis in patient vital signs [25], but to the best of our knowledge researchers have not trained ML algorithms only on biosignals from bedside monitors to differentiate patient age and gender. This is a necessary analysis, and offers insights into whether differences in biosignals can unintendedly be learned by a model in a discriminatory way, and therefore make the model predict in favour of certain patient populations at the expense of others.

The paper is organised as follows. Section two describes the data used for the research as well as the applied methodology. Section three contains the results and discussion, whilst section four concludes the paper.

2 METHODOLOGY

This section outlines the data used for the research and the specifics of the preprocessing stage. Additionally, we give an overview of the algorithms used as well as the metrics which evaluate the trained models.

For the purposes of this research, we use the VitalDB dataset [26], which contains biosignals and clinical information from 6,388 non-cardiac surgical patients that underwent surgery in Seoul National

University Hospital in Seoul, Republic of Korea. The data has high-resolution with 2.8 million data points per case on average. The data of interest for us included: from demographic information, age and gender, and from vital information measured using Solar 8000M monitor, heart rate, respiratory rate, and (systolic, diastolic, and mean) blood pressure both invasively and non-invasively measured. As each of the biosignals was organised in a separate file, before proceeding with training the algorithms, it was necessary to merge the information while minding the time stamp of each measurement in order to maintain the continuity of the data. Additional information related to the surgical approach and the anaesthesia were not considered. The selected data was analysed in two different formats: first, the original data as recorded by the monitor without interference, and second, using features obtained with the tsfresh library [27]. In both cases, only patients with measurements for all biosignals of interest were considered, which reduced the population to 2905 patients.

The analysis of the demographic information, age and gender, is separate; namely, the gender analysis is a binary classification, whereas the age analysis is a multiclass classification problem. Each analysis was conducted both with the original data and with tsfresh statistics from the original data. All cases consider several ML algorithms: Logistic Regression (LR) [28], which estimates the probability of an event occurring, and so establishes baseline results, then Decision Trees (DT) [29] which represents a tree-like model showing series of decisions and possible consequences, Random Forest (RF) [30] which contains a collection of trees and uses a majority voting system to obtain the final prediction, and XGBoost [31], which compared to Random Forest operates on adjustable parameters through iterations, is proven as the most successful algorithm, even in cases of small and medium datasets, with limited feature count, as is the case here. However, as XGBoost is prone to overfitting when trained on small data, we performed parameter optimisation so to restrict the expansion of the model's structure.

The evaluation of the classification for each of the models was performed using a confusion matrix and a classification report (which observes metrics across each class), both for binary and multiclass classification. The confusion matrix visually represents the performance of the models, as it summarises the predicted and actual values obtained from the model and illustrates all misclassifications. The classification report shows the performance for each individual class and provides overall metrics for all classes. It observes the overall accuracy of the model and provides precision, recall, and F1-score values for each class. Precision measures how many of the positive predictions made are in fact correct, whilst recall measures how many of the positive cases from the overall positives were correctly predicted. The F1-score combines both metrics and shows intel into how many times the model made a correct prediction across the entire dataset.

3 RESULTS

The obtained results are divided into two separate groups: binary classification results for gender bias and multiclass classification results for age bias. The age bias results observe two age range divisions: one in three groups and another in four groups. The division of age ranges in three subgroups resulted in the first group of patients under 30 years, the second with patients between 30 and 49 years, and the third contained patients aged 50 and above. As majority of patients were aged 50 and over, and considering the age range considered for the third group was larger, we extended the analysis into a division of four groups, where the third range was split in two, with patients aged 50 to 69 years, and another with patients aged 70 and above.

3.1 Gender Bias

In order to perceive gender bias, the biosignals are used to classify patients as either male or female. The accuracy for all algorithms, both trained on the original data and the tsfresh features given in Table 1.

Table 1: Accuracy from gender bias analysis.

Models	Original Data	TSFRESH Features
LR	64%	61%
DT	99%	53%
RF	100%	63%
XGBoost	84%	58%

As can be observed from the Table 1, the prediction is better when trained on the original values of the data. As expected LR provides the baseline result, whereas the three remaining algorithms show improvement in performance. The accuracy of 84% for XGBoost shows that gender can be identified from biosignals in four from five patients, which is a significant number. The two remaining algorithms show an accuracy of 99% and 100% respectively, which essentially indicates that

biosignals can help AI algorithms to identify all patients' gender details.

The precision, recall, and F1-score are structured in Table 2. With XGBoost exists a drop in predictive power between the two classes, which is not the case with the results from DT and RF. The drop in the metrics for female patients can partially be due to a smaller pool of female patients. Nevertheless, these results are consistent with previous research data showing male patients have higher blood pressure compared to females [32].

Table 2: Classification report from gender bias analysis on the original data (M – male, F – female) (in %).

Models	Precision		Recall		F1-score	
	M	F	M	F	M	F
LR	64	61	87	29	74	39
DT	99	99	99	99	99	99
RF	100	100	100	99	100	100
XGBoost	83	85	91	73	87	78

This shows that models are able to detect subtle differences in data between patients of different genders, and while these subtle differences are necessary when analysing blood pressure information, they are not a beneficial feature when analysing biosignals in general, since models' performances need to be invariant to demographic information.

3.2 Age Bias

The results for age bias, obtained using the selected biosignals, are observed from two standpoints: first, where only three groups of patients are considered, and second, with four groups of patients considered (created by dividing one of the three groups from the first observation into two). As this approach uses multiclass classification, only three algorithms were considered; namely LR was not trained and tested for these data points. The results from the division of patients in three groups (under 30; between 30 and 49; 50 and over) are given in Table 3. The results from the division of patients in four groups (under 30; between 30 and 49; between 50 and 69; 70 and over) are given in Table 4.

Table 3: Accuracy from age bias analysis (3 groups).

Models	Original Data	TSFRESH Features
DT	99%	69%
RF	100%	76%
XGBoost	91%	75%

These results show that patients' age groups can be identified using biosignal information with an accuracy of 100% when using RF. The high accuracy results are obtained on the original data without value interference, whereas processing the data and using features extracted with tsfresh results in significant decrease of performance. When observing the behaviour of the models on the train and test data, the differences in metrics indicate that the models overfit when trained on the tsfresh features, which partially accounts for the worsened performance. Another reason is the difference in data points, meaning as there is lower data point count with tsfresh (since this approach aggregates the original data) the model is impacted by that reduction.

Table 4: Accuracy from age bias analysis (4 groups).

Models	Original Data	TSFRESH Features
DT	98%	43%
RF	99%	54%
XGBoost	80%	51%

With DT and RF obtaining near perfect results, it is interesting to analyse the performance of XGBoost and potential reasons for its performance. The confusion matrix from the analysis of three groups using the original data, given in Figure 1, shows the model mistakes patients aged between 30 and 49 with patients aged 50 and over, which might indicate that the model struggles with differentiating blood pressure values per age [33]. Potential conflicts in age-related medical problems can stem from differences in biological and chronological age [34], however with the other two algorithms performing with an accuracy approaching 100%, this is unlikely the case here.

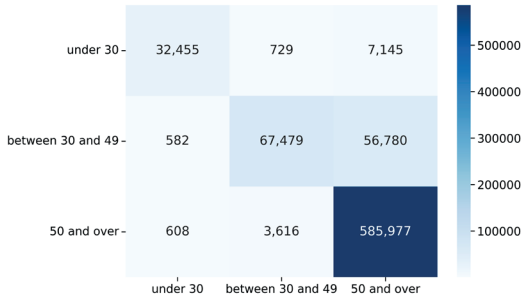


Figure 1: Confusion matrix for original data with three age groups analysed.

Another thing which can be noted is that the performance of the models decreases when patients aged 50 and over are divided in two groups, with DT dropping from 99% to 98%, RF dropping

from 100% to 99%, and XGBoost significantly dropping from 91% to 80%. The change in performance can be observed in the confusion matrix for the original data for four groups, as seen in Figure 2. Namely, once the patients are divided, the model is impacted and unable to successfully learn the difference between patients aged 50 to 69 and patients aged over 70. As the confusion matrix shows, a third of patients aged over 70 are misclassified into the group containing patients aged 50 to 69.

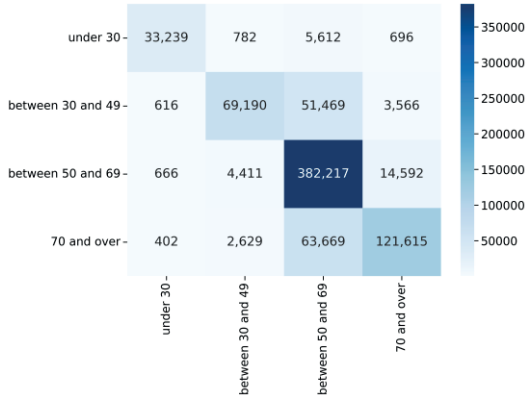


Figure 2: Confusion matrix for original data with four age groups analysed.

3.3 Discussion

The results for both gender and age bias show that ML algorithms are able to differentiate genders and age ranges based on biosignals, which in turn shows that identifying potential biases in data can be accomplished by observing whether specific input information can be used to predict classes belonging to a variable carrying said potential bias. In cases where the algorithms accomplish near perfect score, as is the situation here with DT and RF, it is safe to say that using the data in the same format might confuse the algorithms and lead them to predict based on information which should be disregarded.

With results showing high accuracy in predicting demographic information, it is necessary to discuss potential reasons behind the successful performance of the models. Namely, differences in biosignals based on gender and age have been shown, and it is likely the ML models observe these differences and make predictions on them. With models being able to differ between patient groups based on biosignals, it is possible that ML models trained on these biosignals for various other medical purposes also make their decisions based on these differences, and adjust predictions based on demographic information.

Therefore, another interesting discussion to touch up on are the implications of these results and the challenges which they pose for real-world use of ML algorithms in medical setting. Namely, implicit bias can easily be propagated along the pipeline, and create biased application, which in turn can lead to skewed outcomes and inequity among different patient populations. This can lead to favouritism of certain patient groups as well as reduced or inaccurate performance of models based on demographics. Depending on the application and the purpose of the algorithms, serious illnesses can be disregarded or overlooked, patients can be silenced on important health problems, patients might receive substandard preventive care, and many others. All of the above can lead to higher chances of worsening medical conditions, health complications, disruption of patients' lives, and in extreme cases, deaths which could have been avoided.

4 CONCLUSIONS

This paper proposes a demographic bias analysis approach from patients' biosignals, using ML algorithms to perform binary and multiclass classification in order to identify patient gender and age. The approach focused on analysing two types of results, firstly, the original data was used, and secondly, the data was processed and extracted tsfresh features were used. In both cases, bias could be seen, however bias was more prominent with the original data. This indicates that extracting features using tsfresh can be seen as a marginal mitigation technique in partially handling bias in this dataset. However, further research is required in order to understand whether the same holds for other data. Moreover, with results showing that biosignal information can be used to classify patients according to gender and age (with two separate analyses into three and four age groups), the approach can allow researchers to understand whether algorithms might detect hidden bias in data which cannot be easily observed by the developer. Therefore, the approach itself can be used to mitigate potential biases in creating and selecting datasets, as well as throughout the processing stages when developing AI-based medical systems. This would reduce the propagation of biased data and practices in real-world applications before they are deployed into production, which would greatly benefit patients discriminated upon by biased applications.

REFERENCES

[1] J.A. Sabin, "Tackling implicit bias in health care," N. Engl. J. Med. 2022, vol. 387, pp. 105-107, July 2022.

[2] J.R. Marcelin, D.S. Siraj, R. Victor, S. Kotadia, and Y.A. Maldonado, "The impact of unconscious bias in healthcare: how to recognize and mitigate it," The Journal of Infectious Diseases, vol. 220, issue Supplement_2, pp. S62-S73, September 2019.

[3] J.C. Lester, S.C. Taylor, and M.M. Chren, "Under-representation of skin of colour in dermatology images: not just an educational issue," British Journal of Dermatology, vol. 180, no. 6, pp. 1521-1522, June 2019.

[4] American Cancer Society, "Cancer Facts and Figures 2023,". [Online]. Available: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2023/2023-cancer-facts-and-figures.pdf, [Accessed on Sep. 24, 2023].

[5] J.F. Dovidio, S. Eggly, T.L. Albrecht, N. Hagiwara, and L.A. Penner, "Racial biases in medicine and healthcare disparties," TPM, vol. 23, no. 4, pp. 489-510, December 2016.

[6] L. VanPuymbrouck, C. Friedman, and H.A. Feldner, "Explicit and implicit disability attitudes of healthcare providers," Rehabilitation Psychology, vol. 65, no. 2, February 2020.

[7] A. Dugravot, A. Fayosse, J. Dumurgier, K. Bouillon, T.B. Rayana, A. Schnitzler, Prof. M. Kivimaki, S. Sabia, and Prof. A. Singh-Manoux, "Social inequalities in multimorbidity, frailty, disability, and transitions to mortality: a 24-year follow-up of the Whitehall II cohort study," The Lancet Public Health, vol. 5, issue 1, pp. E42-E50, January 2020.

[8] H.R. Burstin, S.R. Lipsitz, and T.A. Brennan, "Socioeconomic status and risk for substandard medical care," JAMA., vol. 268, issue 17, pp. 2383-7, November 1992.

[9] A. Samulowitz, I. Gremyr, E. Eriksson, and G. Hensing, ""Brave men" and "emotional women": A theory-guided literature review on gender bias in health care and gendered norms towards patients with chronic pain," Pain Res. Manag., vol. 2018, pp. 6358624, February 2018.

[10] D. Westergaard, P. Moseley, F.K. Hemmingsen Sørup , P. Baldi, and S. Brunak, "Population-wide analysis of differences in disease progression patterns in men and women," Nat. Commun., vol. 10, no. 1, pp. 666, February 2019.

[11] B.N. Greenwood, S. Carnahan, and L. Huang, "Patient–physician gender concordance and increased mortality among female heart attack patients," Proc. Natl. Acad. Sci. U S A, vol. 115, no. 34, pp. 8569-8574, August 2018.

[12] A. Ben-Harush, S. Shiovitz-Ezra, I. Doron, S. Alon, A. Leibovitz, H. Golander, Y. Haron, and L. Ayalon, "Ageism among physicians, nurses, and social workers: findings from a qualitative study," Eur. J. Ageing, vol. 14, no. 1, pp. 39-48, March 2017.

[13] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," Science, vol. 366, no. 6464, pp. 447-453, October 2019.

[14] L. Seyyed-Kalantari, G. Liu, M. McDermott, I.Y. Chen, and M. Ghassemi, "CheXclusion: Fairness gaps in deep chest X-ray classifiers," Pac. Symp. Biocomput., vol. 26, pp. 232-243, 2021.

[15] R. Lutz, Incident Number 110. In McGregor, S. (ed.) Artificial Intelligence Incident Database, Responsible AI Collaborative, 2023.

[16] A.J. Larrazabal, N. Nieto, V. Peterson, D.H. Milone, and E. Ferrante, "Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis," Proc. Natl. Acad. Sci. U S A, vol. 117, no. 23, pp. 12592-12594, May 2020.

[17] Sklearn, "Classification report,". [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html, [Accessed on Sep. 24, 2023].

[18] Sklearn, "Confusion matrix,". [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html?fbclid=IwAR2TABvX6ymzxMD1Db7cV-wzNOFI6WzifcZHQez4FrVj0f6iCWv9RZON8es, [Accessed on Sep. 24, 2023].

[19] K. Prabhavathi, K.T. Selvi, K.N. Poornima, and A. Sarvanan, "Role of Biological Sex in Normal Cardiac Function and in its Disease Outcome – A Review," Journal of Clinical and Diagnostic Research, vol. 8, no. 8, pp. BE01–BE04, August 2014.

[20] A. LoMauro, and A. Aliverti, "Sex differences in respiratory function," Breathe (Sheff), vol. 14, no. 2, pp. 131-140, June 2018.

[21] R. Maranon, and J.F. Reckelhoff, "Sex and Gender Differences in Control of Blood Pressure," Clinical Science, vol. 125, no.7, pp. 311-318, October 2013.

[22] C.E. Boeke, M.E. Pauly, H.H. Stock, S. Pavlis, and J. B. Jackson, "The association of gender, age, body mass index, and vital signs in healthy plateletapheresis donors," Transfusion and Apheresis Science, vol. 41, no.3, pp. 175-178, December 2009.

[23] M. Erbilek, M. Fairhurst, and M.C.D.C. Abreu, "Age Prediction from Iris Biometrics," IET, December 2013.

[24] M. Rajput, and G. Sable, "Deep Learning Based Gender and Age Estimation from Human Iris,", SSRN Electronic Journal, June 2019.

[25] B. Velichkovska, H. Gjoreski, D. Denkovski, M. Kalendar, B. Mamandipoor, L.A. Celi, and V. Osmani, "Vital signs as a source of racial bias," medRxiv, February 2022.

[26] H.C. Lee, Y. Park, S.B. Yoon, S.M. Yang, D. Park, and C.W. Jung, "VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients," Sci. Data, vol. 9, no. 1, pp. 279, June 2022.

[27] Tsfresh, "tsfresh,". [Online]. Available: https://tsfresh.readthedocs.io/en/latest/, [Accessed on Sep. 24, 2023].

[28] Sklearn, "Logistic regression,". [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html, [Accessed on Sep. 24, 2023].

[29] Sklearn, "Decision Trees,". [Online]. Available: https://scikit-learn.org/stable/modules/tree.html#decision-trees, [Accessed on Sep. 24, 2023].

[30] Sklearn, "Random Forest,". [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html, [Accessed on Sep. 24, 2023].

[31] XGBoost, "XGBoost Documentation,". [Online]. Available: https://xgboost.readthedocs.io/en/stable/?fbclid=IwAR0s6F9c8B3SFqZ6hvn81OhByL0e-a00h06g8rnRSUEzrUkYdGqXPd-1Yq4, [Accessed on Sep. 24, 2023].

[32] J.F. Reckelhoff, "Gender differences in the regulation of blood pressure," Hypertension, vol. 37, no. 5, pp. 1199-1208, May 2001.

[33] J.L. Lapum, M. Verkuyl, W. Garcia, O. St-Amant, and An. Tan, Vital sign measurement across the lifespan, 1st Canadian ed., Ryerson University, 2018, pp. 123-124.

[34] M.R. Hamczyk, R.M. Nevado, A. Barettino, V. Fuster, and V. Andrés, "Biological versus chronological aging," J. Am. Coll. Cardiol., vol. 75, no. 8, pp. 919-930, March 2020.

EmoStudent: Developing a Dataset to Analyse Students' Emotional Well-Being

Svitlana Antoshchuk and Anastasiia Breskina
Department of Information Systems, Institute of Computer Systems, Odesa Polytechnic National University, Shevchenko Avenue 1, Odesa, Ukraine
asg@op.edu.ua, anastasiia.breskina@op.edu.ua

Keywords: Dataset, Computer Vision, Emotion Understanding, Artificial Intelligence-Based Proctoring Systems.

Abstract: This article introduces an initial version of a dataset designed to educate and assess models that concentrate on studying of the emotional condition of students throughout the remote learning process. This dataset comprises short video clips showing the faces of individuals from diverse ethnic backgrounds and age groups in front of the computer screen. No dataset specialising in the proctoring systems problem solving task was found (process of working in front of the computer, emotions during the educational process). As a result, existing datasets for solving problems in related fields were analysed: emotion classification, emotion recognition, and face recognition. Building on this analysis and the specifics of the chosen data source (YouTube videos with Creative Commons license), the previously established criteria for creating the dataset were modified and expanded. A more adaptable approach was introduced concerning the categorization based on age and ethnicity. A path for future endeavors was also delineated, proposing an enhancement of the current implementation to encompass a broader spectrum of emotions and individuals with various forms of disabilities in subsequent iterations.

1 INTRODUCTION

Throughout the history of distance learning technologies there has always been a challenge in evaluating students' behavior and integrity during exams and individual assignments. To tackle this issue, proctoring systems were introduced. These proctoring systems serve as information systems aimed at supervising test or exam completion and monitoring and assessing students' honesty. Essentially, they take on the role of a teacher by observing and evaluating student conduct.

Initially, synchronous proctoring systems relied on human observers, such as teachers or hired staff, to monitor students in real-time. However, in recent years, with the advancement of artificial intelligence (AI) methods and models, there is a possibility to automate the assessment of student integrity [1, 2].

However, students encountered a lot of problems with these systems [1, 3]. Especially the computer vision modules were the subject of many complaints.

The problem was that the systems both had excessive requirements for students' behavior (e.g., the requirement not to take their eyes off the monitor for more than a certain number of seconds) and incorrectly detected the activity (e.g., students of different ethnicities). To address these problems, new rules for online proctoring systems and requirements for student behavior during online testing were proposed [3] to make the systems more humanoriented. More specifically, it was proposed to implement a complex analysis of student behavior, part of which is the analysis of the student's emotional state.

This paper focuses on datasets that are used to train and evaluate machine learning models that used for tasks such as emotion classification, emotion recognition and face recognition. Particularly, it delves into the implementation of an initial implementation of a dataset designed to train and evaluate models that are focused on analysing the emotional state of students during the distance learning process.

2 LITERATURE REVIEW

The task of face recognition is not a new one. There are many datasets created to train and evaluate models that solve these problems. These datasets can be divided into two broad types: image datasets and video datasets.

The most numerous of these are datasets consisting of a set of images of people or just their faces. They are mainly used for face recognition tasks and are characterized by large and high-quality sampling. Some of the well-known datasets relevant to face recognition task are as follows:

- COCO-WholeBody;
- WIDER FACE;
- FDDB (the Face Detection Dataset and Benchmark);
- AFW (Annotated Faces in the Wild);
- PASCAL FACE.

COCO-WholeBody [4] is a modified version of the COCO dataset, enriched with whole body annotations. This enhanced dataset includes four distinct types of bounding boxes for each person depicted in the images: face frame, upper body frame, left arm frame, and right arm frame. Additionally, it provides a substantial number of key points – 133 in total-comprising 17 points for the body, 6 points for the legs, 68 points for the face, and 42 points for the arms.

The WIDER FACE dataset [5] is used for face recognition is derived from the publicly accessible WIDER dataset. It comprises a total of 32,203 images and includes a large number of dummy 393,703 faces. The dataset exhibits significant diversity in terms of scale, pose, and occlusion, presenting a wide range of challenges for face recognition algorithms.

FDDB [6] is a dataset comprising labeled faces extracted from the Faces in the Wild dataset. It contains a total of 5171 face annotations, with images having various resolutions such as 363x450 and 229x410. The dataset encompasses a wide spectrum of challenges, including faces captured at difficult angles, defocused faces, and low-resolution images. Notably, both grayscale and color images are present in this dataset, offering a diverse and comprehensive set of face detection scenarios for evaluation and benchmarking purposes.

AFW [7] is a specific face detection dataset comprising 205 images that collectively contain 468 annotated faces.

The PASCAL FACE dataset [8] is designed for face detection and recognition purposes. It includes 851 images, which are extracted from the larger PASCAL VOC dataset, and encompasses a total of 1341 annotations for faces. However, it's important to note that this dataset is relatively small, containing only a few hundred images, and lacks diverse variations in the appearance of faces. As a result, it may offer limited coverage of real-world scenarios and may not fully represent the wide range of challenges faced in face detection and recognition tasks.

A limitation of these datasets is the lack of a temporal component, which limits the process of analyzing video sequences for human emotional state. To address this problem, video sequence datasets have been created. These include, for example, such sets and benchmarks as:

- RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song);
- AM-FED; CK+ (the Cohn-Kanade database);
- MMI;
- UNBC-McMaster Pain;
- EngageNet.

RAVDESS [9] is a comprehensive dataset containing 7356 files. Within this database, there are recordings from 24 professional actors, equally divided between 12 females and 12 males. The actors provide two lexically comparable utterances in a neutral North American accent.

AM-FED [10] dataset consists of 42 webcam videos that have been recorded in authentic realworld settings. This dataset captures various scenarios and conditions that people commonly encounter when using webcams, making it an ideal resource for researchers and developers interested in real-life webcam-related applications and analysis.

CK+ [11] is a highly popular for advancing facial action and expression recognition systems. Within the CK+ database, there are 593 records of both posed and unposed sequences, meticulously encoded using the Facial Action Coding System (FACS) [12]. Moreover, the sequences are also encoded to represent the six fundamental emotions.

The MMI [13] dataset encompasses a vast collection of face videos, each encoded using the FACS. Specifically, it comprises a total of 1,395 video sequences, with each one manually encoded by identifying the Action Units (AUs) present in the facial expressions;

UNBC-McMaster Pain [14] is considered one of the most extensive collections of AU-encoded video sequences showcasing natural and spontaneous facial expressions. With a total of 10 action units, this

dataset offers a wealth of valuable information. However, it's important to note that even though the expressions are natural and spontaneous, the videos were recorded under controlled conditions..

EngageNet [15] is a dataset comprising a total of 31 hours of recorded data from 127 participants. This dataset captures various individuals under different illumination conditions, providing a diverse and comprehensive representation of facial expressions and interactions.

When it comes to datasets created specifically for the task of face recognition and emotional state recognition of students, there are even fewer of them. Such datasets usually consist of image sets. For example: Face dataset by Generated Photos (Face dataset for Academics by Generated Photos); the CUHK Face Sketch (CUFS).

Face dataset by Generated Photos [16] consists of 10,000 synthetic photos, each carefully balanced with diverse representations of race and gender. It also includes essential metadata and facial landmarks for research purposes. It's important to note that all the photos in this dataset are entirely synthetic, generated based on models, and not real photographs.

CUFS [17] database encompasses a diverse collection of 188 faces from the Chinese University of Hong Kong (CUHK) student database, along with 123 faces from the AR database and 295 faces from the XM2VTS database.

The analysis of existing solutions showed that existing datasets for face recognition and emotion analysis are not suitable for proctoring task for several reasons. Firstly, the video datasets represented people of limited age and ethnicity (usually young adults of Northern European or Asian appearance). Secondly, datasets with only images lose the temporal component. Some of the datasets consist of recordings of professional actors' reactions rather than emotions in the wild;

3 EMOSTUDENT DATASET

The aim of the paper is to develop the first iteration of a dataset that contains videos with emotions in the wild for training and evaluating a model of analyzing students' emotional state during distance learning.

During the conceptualization of the architecture of the automated online proctoring system [3], a prototype dataset was implemented, which is oriented specifically for solving problems within distance learning and online proctoring systems. Further work in this area revealed a number of limitations that led to the refinement of the dataset requirements.

The dataset was sourced from YouTube, specifically from videos that possess a Creative Commons license. The use of this license does not guarantee that there will be non-professional actors in the video. However, in order to get as close to the emotions "in the wild" as possible, it was decided to choose videos with interviews and amateur bloggers. To simplify the labelling process and facilitate dataset creation, the Amazon SageMaker platform was used.

3.1 Requirements for Data

To address the challenge of accurately determining the age of individuals in video data, a different approach was proposed. Instead of categorizing the sample based on specific age intervals, the division was made into broader groups: teenagers, young adults, and adults. This decision was taken due to the difficulty of precisely identifying individual ages from video footage.

The minimum age of children requirement remains the same due to the regulations and policies of various Internet platforms, which often impose a minimum age restriction of 13 years to ensure child safety and compliance with legal guidelines.

To formulate the proportions by age, it is proposed to take as a benchmark the data from the Education Data Initiative project, which collects data and statistics on the education system in the United States. Base on that data 66% of students are under 24 years old, 22.9% are young adults and 10.5% are adults.

The problem of obtaining data depending on people's ethnicity has also arisen. This is because people from different parts of the world have different levels of accessibility to Internet connection and portable devices and cameras. This caused a problem with detailed representation of different nationalities and ethnicities.

For simplicity, it was chosen to divide the data by ethnicity and region of residence: Africa, Caribbean, Asia-East, Asia-South, Europe-North, EuropeCentral, Europe-South, Latin America, Caucasus, Middle East, and Natives.

In the context of separation by region of residence, it is not recommended to form any dependency on the distribution of real data, as it is irrelevant in the context of solving the proctoring task. Instead, it is suggested to form a sample with an equal number of videos to represent each region.

Representation of people of different sex is also suggested to be done with equal representation.

To evaluate student engagement during the learning process, modified 0-3 grading scale designed for assessing student mindfulness [18] was used.

A score of 0 on the engagement scale denotes complete disengagement, where the subject exhibits minimal interest, frequently diverts their attention away from the screen, and may engage in active physical activities. Represented by emotions such as fear, overexcitement, anger. This rating indicates a lack of involvement or attentiveness during the observed activity.

A rating of 1 suggests minimal engagement, indicating that the subject shows only slight interest or involvement. They may barely open their eyes, look at the screen and display restlessness in their chair, suggesting a low level of. Represented by emotions such as bored.

It is proposed to change the essence of the score 2. It will indicate that the subjects are engaged in the content, show interest and interact with the proctoring system, but are in a stressful state. This includes negative emotions (anger, fear, sadness, disgust) combined with direct gaze at the monitor;

A rating of 3 on the engagement scale represents a high level of engagement. This indicates a strong sense of attentiveness and active involvement in the observed activity, showing a clear indication of the subject's deep interest and concentration. Represented by emotions such as neutral and joyful.

3.2 Characteristics of the Developed Dataset

Based on that final requirements for dataset include people from 11 regions; people of different age groups: teenagers, youth and adults. different type of engagement assessment, each of which is respectively described by a different type of student's emotional state.

In this dataset, there were originally 334 video files, each with a duration ranging from one minute to two and a half hours (Table 1). All of them have been split into small video sequences of maximum 1 minute length and 740p size. Average video sample length is 20 seconds; minimum number of the original video splits is from 1 to 50. This resulted in a total of 2816 videos.

The video sequences in this dataset capture individuals' facial expressions, with each recording featuring one person on stage. These individuals exhibit a range of emotional states, including basic emotions such as joy, neutral emotions, anger, sadness, and overexcitement. The dataset provides valuable insight into how these various emotions are expressed through facial cues and expressions, making it a valuable resource for studying and analyzing human emotions in different contexts “in the wild”.

For the labeling process, a group of three annotators was tasked with viewing and assessing the videos, paying particular attention to the level of engagement depicted through facial expressions. This approach allowed for a comprehensive evaluation of the subjects' attentiveness and involvement during the videos, as reflected in their facial expressions and movements. In the created dataset, the distribution of scores is approximately 30% zero mark, 10% one mark, 25% and 35% are two and three marks respectively.

Table 1: Original set of video sequences, categorized by ethnicity and age of people.

Ethnicity (location)/Age	Teen	Young Adult	Adult
Africa	9	10	11
Caribbean	3	10	12
Asia-East	20	25	35
Asia-South	7	4	7
EuropeanNorth	12	9	17
EuropeanSouth	10	10	13
EuropeanCentral	5	2	10
Latin America	5	5	13
Caucasus	7	10	12
Middle East	7	7	14
Natives	1	2	10
Total	86	94	154

The video sequences in this dataset capture individuals' facial expressions, with each recording featuring one person on stage. These individuals exhibit a range of emotional states, including basic emotions such as joy, neutral emotions, anger, sadness, and overexcitement. The dataset provides valuable insight into how these various emotions are expressed through facial cues and expressions, making it a valuable resource for studying and analyzing human emotions in different contexts “in the wild”.

For the labeling process, a group of three annotators was tasked with viewing and assessing the videos, paying particular attention to the level of engagement depicted through facial expressions. This approach allowed for a comprehensive evaluation of

the subjects' attentiveness and involvement during the videos, as reflected in their facial expressions and movements. In the created dataset, the distribution of scores is approximately 30% zero mark, 10% one mark, 25% and 35% are two and three marks respectively.

The audio for the videos was disabled to avoid any auditory influence since in the obtained sample the video sometime contains ambient sounds rather than the speech of the person on the video themselves.

3.3 Current Challenges and Further Plans

The existing dataset implementation faces a few challenges that need to be addressed. One of the issues is the heterogeneous nature of the data, especially concerning shot quality. The videos have been rescaled to a standard 740p size, but some of them still have lower quality compared to others.

Additionally, although efforts were made to include people from various ethnic backgrounds, there is currently an underrepresentation of certain ethnic groups, specifically "Natives" and "European" (especially "European-Central"). There is also a lack of representation for individuals with different types of disabilities.

To address these shortcomings and enhance the dataset's comprehensiveness, the next versions of the dataset implementation are planned to rectify these issues. Steps will be taken to increase the representation of individuals with different types of disabilities and expand the coverage of different ethnicities. These improvements aim to create a more inclusive and comprehensive dataset that better reflects real-world scenarios.

4 CONCLUSIONS

In this research paper, an initial dataset implementation designed to train and assess models with a focus on analyzing the emotional state of students during distance learning was introduced, also several most widely known datasets were reviewed. A notable aspect of this dataset is its inclusion of short video sequences capturing the faces of individuals from diverse ethnic backgrounds and various age groups, all within natural environments.

To create this dataset, a thorough analysis of various existing datasets commonly used for tasks like emotion classification, emotion recognition, and face recognition was made. Given the specific of the chosen data source (YouTube videos with Creative Commons licenses) and its limitations, certain adaptations and expansions were implemented to the existing criteria for dataset creation. Notably, a more adaptable approach was introduced when categorizing data based on age and ethnicity.

This study also presents potential avenues for future research. Specifically, efforts should be made to ensure a balanced representation of individuals with different ethnic background. Additionally, in forthcoming iterations of the dataset, it is planned to include data about individuals with different types of disabilities that can impact their emotional well-being evaluation. By including a wider range of people from different backgrounds, the dataset can contribute to a more accurate and in-depth understanding of emotional reactions in distance learning.

REFERENCES

[1] A. Nigam, R. Pasricha, T. Singh, and P. Churi, "A systematic review on AI-based proctoring systems: Past, present and future," Education and Information Technologies, vol. 26, pp. 6421-6445, Dec. 2021, doi: 10.1007/s10639-021-10597-x.

[2] S. Motwani, C. Nagpal, M. Motwani, N. Nagdev, and A. Yeole, "AI-Based proctoring system for online tests," Proceedings of the 4th International Conference on Advances in Science & Technology (ICAST2021), Jun. 2021, doi: 10.2139/ssrn.3866446.

[3] A. Breskina, "Development of an automated online proctoring system," Herald of Advanced Information Technology, vol. 6, no. 2, pp. 163-173, 2023, doi: 10.15276/hait.06.2023.11.

[4] S. Jin et al., "Whole-Body human pose estimation in the wild," ECCV 2020: Computer Vision – ECCV, 2020, pp. 196-214, 2020, doi: 10.48550/arXiv.2007.11858.

[5] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A face detection benchmark," 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 5525-5533, 2016, doi: 10.1109/CVPR.2016.596.

[6] V. Jain and E. Learned-Miller, "FDDB: A benchmark for face detection in unconstrained settings.," Technical Report UM-CS-2010-009, Dept. Of Computer Science, University of Massachusetts, Amherst, 2010.

[7] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, pp. 2879-2886, 2012, doi: 10.1109/CVPR.2012.6248014.

[8] J. Yan, X. Zhang, Z. Lei, and S. Z. Li, "Face detection by structural models," Image and Vision Computing, vol. 32, no. 10, pp. 790-799, 2014, doi: 10.1016/j.imavis.2013.12.004.

[9] S. R. Livingstone and F. A. Russo, "The ryerson audiovisual database of emotional speech and song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in north american english," PLoS ONE, vol. 13, no. 5, May 2018, doi: 10.1371/journal.pone.0196391.

[10] D. McDuff, R. E. Kaliouby, T. Senechal, M. Amr, J. F. Cohn, and R. Picard, "Affectiva-MIT facial expression dataset (AM-FED): Naturalistic and spontaneous facial expressions collected "in-thewild"," 2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops, OR, USA,, pp. 881-888, 2013, doi: 10.1109/CVPRW.2013.130.

[11] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended CohnKanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, San Francisco, CA, USA, pp. 94-101, 2010, doi: 10.1109/CVPRW.2010.5543262.

[12] E. A. Clark et al., "The facial action coding system for characterization of human affective response to consumer product-based stimuli: A systematic review," Front Psychol, May 2020, doi: 10.3389/fpsyg.2020.00920.

[13] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," 2005 IEEE International Conference on Multimedia and Expo, Amsterdam, Netherlands, pp. 5, 2005, doi: 10.1109/ICME.2005.1521424.

[14] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews, "Painful data: The UNBC-McMaster shoulder pain expression archive database," 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG), Santa Barbara, CA, USA, pp. 57-64, 2011, doi: 10.1109/FG.2011.5771462.

[15] M. Singh, X. Hoque, D. Zeng, Y. Wang, K. Ikeda, and A. Dhall, "Do I have your attention: A large scale engagement prediction dataset and baselines," 2023. doi: 10.48550/arXiv.2302.00431.

[16] "Face dataset by generated photos (face dataset for academics by generated photos)," [Online]. Available: https://generated.photos/ datasets#research-dataset, [Accessed Jul. 3, 2023].

[17] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," CVPR 2011, Colorado Springs, CO, USA, pp. 513-520, 2011, doi: 10.1109/CVPR.2011.5995324.

[18] A. Kaur, A. Mustafa, L. Mehta, and A. Dhall, "Prediction and localization of student engagement in the wild," 2018 Digital Image Computing: Techniques and Applications (DICTA), Canberra, ACT, Australia, p. 1-8, 2018, doi: 10.1109/DICTA.2018.8615851.

Dynamic Topic Modelling of Online Discussions on the Russian War in Ukraine

Taras Ustyianovych[1], Nadiia Kasianchuk[2,3], Halina Falfushynska[4,6,7], Solomiia Fedushko[5,8] and Eduard Siemens[6]

[1]*Department of Artificial Intelligence Systems, Lviv Polytechnic National University, Kniazia Romana Str. 5, Lviv, Ukraine*
[2]*Faculty of Biology, Adam Mickiewicz University, Uniwersytetu Poznańskiego Str. 6, Poznań, Poland*
[3]*Faculty of Pharmacy, Bogomolets National Medical University, Taras Shevchenko Str. 13, Kyiv, Ukraine*
[4]*Department of Marine Biology, Institute for Biological Sciences, University of Rostock, Albert Einstein Str. 3, Rostock, Germany*
[5]*Department of Social Communication and Information Activities, Lviv Polytechnic National University, Kniazia Romana Str. 5, Lviv, Ukraine*
[6]*Department of Electrical, Mechanical and Industrial Engineering, Anhalt University for Applied Sciences, Bernburger Str. 55, Köthen, Germany*
[7]*Statistical Analysis Center, Wroclaw Medical University, K. Marcinkowski Str. 2-6, Wroclaw, Poland*
[8]*Department of Information Systems, Faculty of Management, Comenius University in Bratislava, Odbojárov Str. 10, Bratislava, Slovakia*
taras.o.ustyianovych@lpnu.ua, nadkas2@st.amu.edu.pl, halina.falfushynska@uni-rostock.de, solomiia.s.fedushko@lpnu.ua, eduard.siemens@hs-anhalt.de

Keywords: Machine Learning Operations (MLOps), Social Media Discussions, Russian War in Ukraine, Splunk Enterprise, Latent Dirichlet Allocation (LDA).

Abstract: The availability of robust end-to-end ML processes plays a crucial role in delivering an accurate and reliable system for real-time text data inference. In this paper, we present an approach to building machine learning operations (MLOps) and an observability application to perform topic modelling of online discussions in social media, here observed based on topics and threads related to the Russian war in Ukraine. Splunk Enterprise is the main tool and platform used throughout this research with its knowledge discovery, dashboarding, and alerting. 30GB of social media text data coming from a Russian social network VKontakte over the time line January 2022 to May 2023. Main inquiries included text mining and topic modelling, which we managed to perform over the observation period using Python frameworks, mainly gensim for text processing and MLflow for experiment management and logging. The Splunk architecture allowed us to ingest and analyse the results and prediction of ML experiments for dynamic topic modelling, and served as a MLOps solution. The designed set of five dashboards played a crucial role in determining the optimal model hyperparameters (number of topics, A-priori belief on document-topic distribution, number of total corpus passes) and drift detection which occurred almost every two-three weeks depending on the phase of the war. Our application assisted us with text analysis, discovering how events on the battlefield influenced social media discussions, and what post attributes contributed to a high user engagement. With our setup we were able to find out how antiwar hashtags have been used to promote misleading content actually supporting the war against Ukraine. The analysis of the researched discussions shows a trend where usage of adjectives decreased over time since the war has started, whereas an increase for nouns and verbs usage over time. Information distortion has steadily been present in the content leading to bias and misleading data in social media discussions.

1 INTRODUCTION

The rapid proliferation of social media platforms has transformed the way information circulates, impacting everything from daily conversations to global events. As these platforms become central hubs of information exchange, they also become fertile grounds for the spreading fakes, so called trolls activities and targeted propaganda campaigns. The

Russian war in Ukraine as a significant geopolitical event can be used as a major example where, amidst the vast amounts of discussions, opinions, and perspectives across various social media platforms, instances of misinformation, orchestrated facts manipulation and down to bare propaganda campaigns can bias the perception of social media discussion For this, several sources of Russian platforms have been observed [1]. This immense amount of real-time data, which is a mixture of genuine insights and manipulated narratives is more and more used as a decision-making base by policymakers, managers of businesses and even by researchers. Analyzing the data in real-time is required to find trends, categorize information into meaningful pieces, map inter-dependencies, and model social media discussions, which can be efficiently accomplished using machine learning capabilities. Therefore, processing raw data through computational algorithms would eliminate the toil, and let analysts and researchers focus on knowledge mining of organized information with a drilldown capability to dive into the original data.

However, harnessing this vast volume of data and extracting meaningful insights from it in real-time presents several challenges. Conventional data analysis tools and methodologies often fall short when faced with the dynamic nature of social media content, which is characterized by its heterogeneity, volume, and velocity [2]. Machine Learning Operations (MLOps) - a set of practices, guidelines, and tools - promises to bridge this gap [3]. MLOps, when combined with robust end-to-end machine learning processes, enables the development of systems that can efficiently process, analyse, and reveal insights from massive streams of real-time text data. The implemented MLOps-powered system improved the observability of topic models, assisted with data drifts detection, and determining the frequency for retraining. We found out that model training time increases with the number of topics for the text to be split into. The analyzed metrics used to evaluate the model performance are correlated in such a way: perplexity has a 79.5% correlation with UMass coherence score and 71.8% with CV coherence scores. The correlation between these two coherence scores is just 27.1%. This confirms that decision for the optimal model hyperparameters can be made in conjunction with these three metrics.

According to the CRISP-ML(Q) model [4], effective monitoring of machine learning operations is as vital as the development of the model itself [3]. This is particularly true in dynamic environments with frequent data change and drift is paramountto maintaining high standards for performance and knowledge mining in such settings.

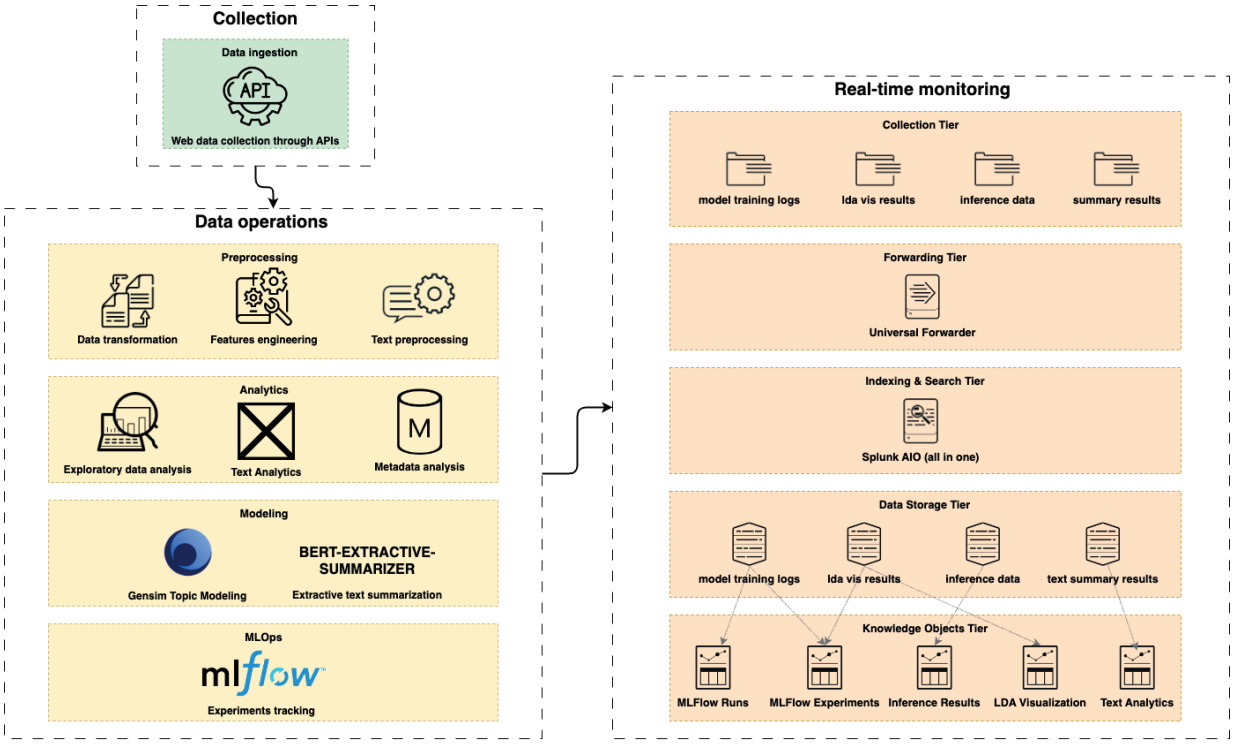


Figure 1: Workflow architecture of the machine learning operations process.

In this paper, we improve the efficiency and user-experience of processing and analyzing vast amounts of social media discussions. Our methodology, underpinned by advanced text mining techniques and topic modelling, not only aids in knowledge discovery but also equips us with the tools to detect model drifts, ensuring the continual relevance and accuracy of our insights. As we navigate through the intricacies of our solution, we also shed light on the importance of real-time monitoring, the challenges posed by the dynamic environment of social media, and the invaluable insights that can be drawn from such an endeavour.

2 MATERIALS AND METHODS

Machine Learning operations and their monitoring are intricately tied to the specific modelling use-case in question. CV and UMass coherence scores and perplexity were used in this research to evaluate the performance of the trained topic models [5]. Similarly, the optimization of hyperparameters varies across algorithms, contingent upon the model architecture. In this context, we delve into the monitoring and evaluation of the *Latent Dirichlet Allocation (LDA)* algorithm, trained in Python using the *gensim* library, to automatically model topics within social media data pertinent to geopolitical events [6, 7]. The LDA algorithm is favoured by numerous researchers due to its efficiency and adaptability. Recent studies underscore its prominence as a generative text model [6]. Given the data's nature, it's imperative to continuously update knowledge in real-time and remain vigilant to a significant number of outliers or any signs of performance degradation. Upon spotting any of these indicators, swift detection and proactive measures are essential for being able to observe and to identivy trends and biases in the subject under study [8]. The presence of robust MLOps becomes even more critical for high-load production systems. Disruptions in service level objectives or agreements (SLOs and SLAs) can have detrimental effects on the system reliability and performance [9].

Having these prerequisites in mind, the architecture depicted in the Figure 1 has appeared conceiving.

According to the architecture diagram (Figure 1), there are three primary layers, namely:

- data collection;
- data operations;
- real-time monitoring.

To understand the use-case and implementation, a concise description of each layer is provided.

Data collection has been performed using officially available APIs from *VKontakte* social media platform to retrieve a discrete number of posts based on a specified query for the period since January 2022 until May 2023. The following keyword parameters were specified in the search query to retrieve the most relevant pieces of information related to the actual Russian war in Ukraine: *"war", "specoperation"*, and *"Ukraine"*. The returned data undergoes a standard procedure involving cleaning, transformations, feature engineering, and text pre-processing. The latter is

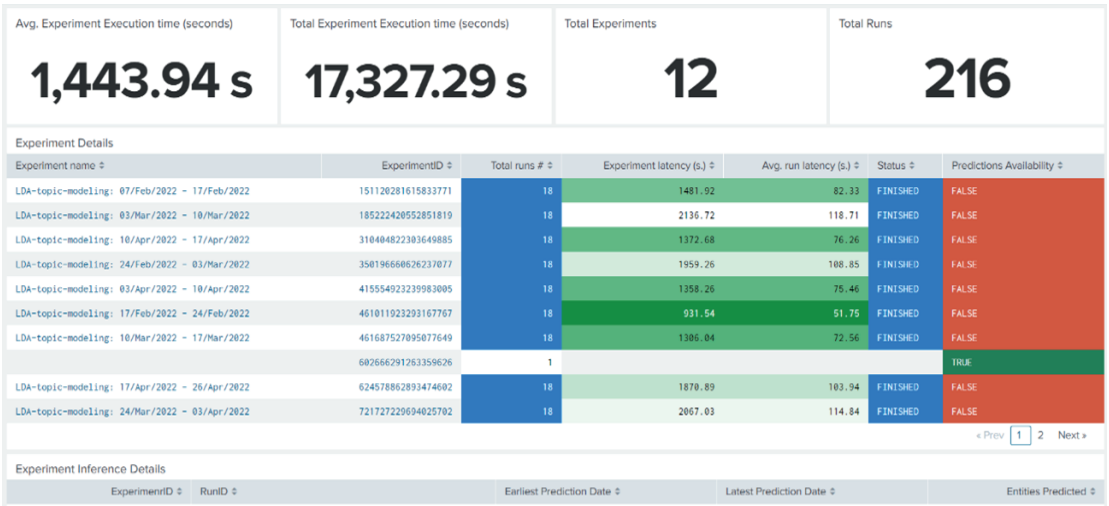


Figure 2: Visual representation of MLFlow experiments, detailing run metrics, execution times, and the relationship between

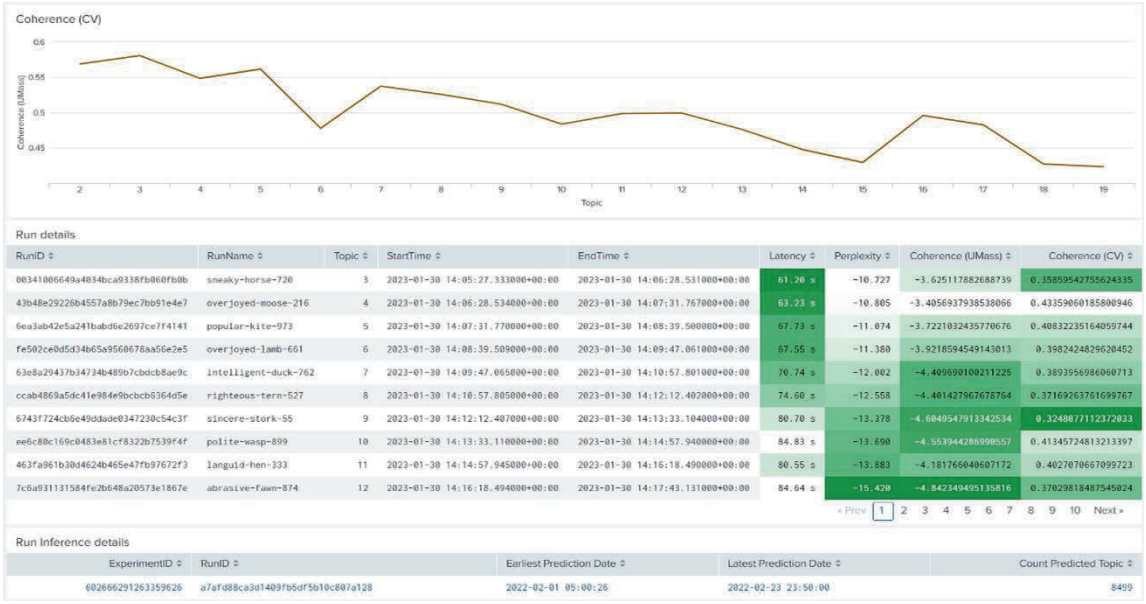


Figure 3: A detailed visualisation of run metrics, emphasizing the correlation between the number of topics and run latency, performance scores, and the exploration of hyperparameters across different runs for optimal model training.

particularly crucial as it serves as input for the LDA model. Subsequently, metadata and consolidated statistics related to the social media posts are stored and passed to the next layer. Text vectorization and model training have been conducted using the *gensim*, *spaCy*, and *NLTK* Python-libraries. For hyper-parameter tuning, experiment management, and subsequent model deployment, the open-source library *MLFlow* has been used.

The monitoring artefacts in the third layer, storing the output from the previous step, are split into four categories, see Figure 1, right part:

- Model training logs that include information about hyperparameters, training duration, experiment and run IDs, metrics, and additional metadata.
- LDA visualization results, which comprise JSON files needed to reconstruct and depict the determined topics of the best-performing model on a 2D Cartesian plot. The outcomes from this category aid decision-making during the model selection process post hyperparameter tuning.
- The inference data category preserves post data, the label assigned by the model, probability scores, and the principal component values of the input documents' vectors.
- Summary results encapsulate text summaries for each identified topic.

Each category contains common keys, such as experiment and run IDs, derived from the preceding step when leveraging MLFlow's logging features.

Once the output data is systematically categorized, the Splunk ecosystem is employed. The monitoring layer's implementation predominantly utilizes Splunk Enterprise, a comprehensive data platform, designed for real-time information indexing, storage, and processing. Splunk was selected due to its developer-centric features and adaptable search processing language.

A universal forwarder, a streamlined version of Splunk Enterprise equipped with data forwarding functionalities, performs real-time data ingestion to a a dedicated Splu nk Enterprise instance tasked with indexing and searching operations. Each index retains data corresponding to its category. For example, model training logs are exclusively ingested into the model training index, while inference data is confined to the inference data index. Leveraging IDs consistent across indexes enables the execution of joined operations and other data-related tasks.

The concluding phase of the third layer encompasses data visualization via interactive dashboards equipped with a variety of dropdowns and filter options. Additionally, for enhanced monitoring, alerts can be configured to detect model performance degradation and identify anomalies.

In summary, five dashboards, termed as "knowledge objects," were formulated within Splunk Enterprise. These dashboards interrelate and

collectively aim to offer real-time insights into MLOps and the intricate analysis of social media data.

3 RESULTS AND DISCUSSIONS

This section describes the derived artifacts, specifi findings their significance along with suggestions for enhancements in future research. We have created the following knowledge objects to aid MLOps and observability implementation:

- Experiments
- Runs
- Topic Modelling Visualization
- Inference Results
- Text Analytics

We will sequentially discuss each of these knowledge objects to elucidate their value, challenges during their development, and inherent limitations.

The Experiments dashboard (Figure 2) showcases information corresponding to a unit within the MLFlow architecture. An experiment comprises multiple runs, where a run is a single execution of a script with defined parameters. In our scenario, the experiment pertains to training a topic modelling model over a specified duration. It's essential to consider hyperparameters like the number of topics, alpha score, and iterations over the dataset to achieve optimal results. Each run encapsulates a specific set of hyperparameters and, upon completion, yields metrics for evaluation and comparison.

The aforementioned dashboard displays summarized statistics about various experiments. It provides filtering options based on time window, experiment name, and ID. The dashboard's single-value panels convey details about the average and total execution time, number of experiments, and runs. The table presents each experiment's comprehensive attributes: its name, unique identifier, number of runs, total and average execution latency per run, status, and an indicator for the availability of prediction results. For each experiment, the optimal run is selected, and its hyperparameters are employed for the final topic modelling. The dashboard contains a table with details about the best run, and if inference results exist, one can navigate to this dashboard to evaluate the topic modelling outcomes by clicking on the selected run.

Evaluating the set hyperparameters of each run within an experiment necessitates comparing performance metrics. The runs dashboard (Figure 3) empowers engineers and analysts to execute this task. Leveraging Splunk's drill-down features, transitioning from Experiments to Runs knowledge objects significantly enhances the user experience. The latter dashboard is prefiltered based on the selected experiment ID from the previous dashboard. Run latency is a crucial consideration during model training. The Runs dashboard displays a bar chart and summary statistics for runs' execution times. A discernible pattern emerges: a higher number of topics correlate with increased run latency. Perplexity and coherence scores are plotted against the number of topics using line charts, facilitating easy correlation of metrics for a given topic count. The same piece of information is replicated below these line charts in a table "Run details" with supplementary metadata about specific runs. The built-in capabilities for table formatting afford a comprehensive view of performance metrics, aiding analysts in their comparison. Furthermore, the dashboard enables the identification of optimal hyperparameters, which can be adjusted in subsequent runs.

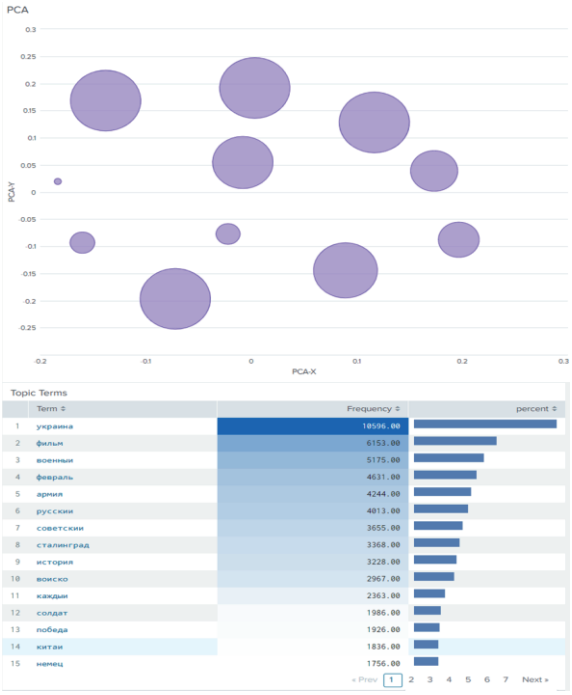


Figure 4: Interactive Topic Modelling Visualization. Displaying modelled topics on a Cartesian plot using the LDA algorithm. The bubble chart demonstrates topic distributions, with size indicating document frequency and proximity suggesting potential topic similarities. The layout aids in discerning term commonalities across topics and highlights the LDA's emphasis on word frequency over sentiment analysis.

After selecting the best runs, the topic modelling model with the most suitable hyperparameters can be trained for inference. Topic modelling objective is to categorize a text dataset into meaningful groups to extract knowledge.

The Topic Modelling Visualization (Figure 4) dashboard provides a suite of panels and filters to:

- Determine the placement of each modelled topic on a Cartesian plot and its relationship to other topics.
- Analyse a topic's keywords and terms.
- Identify common keywords and terms across the modelled topics.

We employed the LDA algorithm for topic modelling. The visual formatting of the panels aids decision-making. For instance, a bubble chart (Figure 4) depicts topics on a Cartesian plot, with larger bubbles indicating topics with more documents. Closely situated or overlapping bubbles might suggest topic similarity, possibly indicating that the topic count is too high. The table with topics terms below shows the most crucial keywords that describe the processed messages. The top terms shown on Figure 4 are “Ukraine”, “movie”, “military” (adjective), “February”, “army”, “Russian”, “soviet”, “Stalingrad”, “history”, “military” (noun), “soldier”, “victory”, “China”, “German”. The frequency and percent columns indicate the quantity of these keywords’ appearance in the input documents. We can conclude most of the collected data based on the input search parameters is related to 1) the war in Ukraine, 2) historical battles during the Second World War, 3) geopolitical situation in the world.

Hence, such visualizations are invaluable, facilitating efficient task completion. Similarly, if specific terms frequently appear in multiple topics, it's a clear sign of their relatedness. This observation highlights an LDA limitation: it doesn't capture sentiment and relies on word counts in the training text data.

The Inference Results dashboard (Figure 5), the fourth one, is designed to showcase predictions made by the model with the best hyperparameters from a given experiment. The data for this dashboard shares common keys, such as experiment and run IDs, enabling linkage with other datasets. Consequently, filters for these IDs are present on the Inference Results dashboard. Primary panels include:

- Document count by topics.
- Average, minimum, and maximum distances to the topics' centroids for the documents.
- Probability box plots to display the likelihood of documents belonging to the predicted topic.
- t-SNE visualization of the predicted document-to-topic associations.
- A table with summary statistics about each topic's distance to its centroid, aiding in identifying topics with numerous outliers.

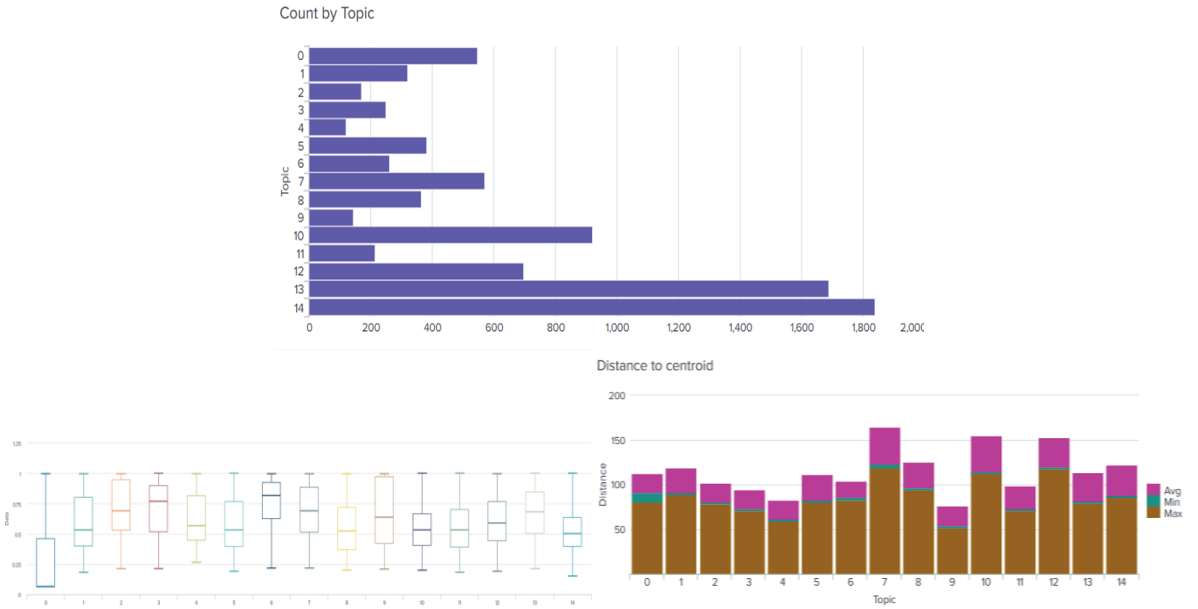


Figure 5: Inference Results Overview. An in-depth display of model predictions based on optimal hyperparameters. This dashboard visualizes topic distributions, document-to-topic distances, and prediction likelihoods. It aids in the real-time evaluation of model performance, detection of outliers, and understanding topic dynamics, offering insights into both broad and specific textual representations within the dataset.

Another objective of the Inference Results dashboard (Figure 5) is to detect the trained model's performance degradation. Indications such as a significant number of outliers or declining probabilities of a document belonging to the predicted topic over time are potent signals of such scenarios. Based on our empirical evaluation, if a topic comprises more than 10% outliers among its documents, the model's performance has diminished, rendering it less adept at predicting the correct topic for unseen documents. Such observations were spotted usually on a regular basis ranging between two to four weeks, and are known as the topic survival time.

We present a categorization of topics from the Inference Results dashboard in Table 1.

We observed that topics with more considerable centroid distances tend to share common terms and keywords. Such topics offer a broader representation of the entire text dataset rather than focusing on specific facets. However, this doesn't render them invaluable. These instances exemplify generic meanings and aid in extracting such text entities. Similarly, the Inference Results panels can assist researchers in determining irrelevant topics and related documents, which might be excluded for other exploration tasks.

Text Analytics, the final dashboard of the developed Splunk application, consists of word cloud visualizations, analyses of extracted hashtags, and named entities. Users can interactively select documents related to particular topics, sorted by engagement metrics like views, likes, shares, and comments.

In this case, the designed dashboards were populated with data from social media discussions about the Russian war in Ukraine. The solution's primary benefit lies in its capability for dynamic topic modelling, swift outlier identification, and signs for model retraining. We monitored how discussion topics and used keywords evolved over time and determined the main patterns. Discussions before the invasion (January-February 2022) can be characterized by such keywords used: “accusations” “threats”, “tension”, “uncertain”. Once the full-scale invasion began, we noticed a drastic increase in the number of generated posts. The main terms related to the Russian war in Ukraine were “betrayal", "surrender", "loss of opportunities". Also, historical figures were frequently mentioned. The third period, after more than a year of war, has seen a slight decrease in adjectives usage. This finding means the sentiment of the messages has tended to become neutral rather than emotionally appellative. We observe more and more nouns in the formed topics during this period. Overall, the comparison of the three periods shows that the war has had a significantly shaped the way social media discussions evolve. The war has also led to a sharp increase in the generated and shared content related to this subject. However, there are some signs that the tone of discussions about the war may be changing, as the conflict drags on. This solution has proven for real-time social media analysis and exploration. The incorporation of Splunk's inherent capabilities and add-ons broadens the visualization range and tools for data interaction and visualization.

Table 1. Categorization of the topics indices introduced at the Inference Results dashboard.

Category	Topics	Mined knowledge
Russian War in Ukraine	4, 9, 11, 14.	These topics have the highest number of documents mainly because of the queries specified during data collection. The distance to centroids is low indicating minimum outliers. This simplifies further detection of related documents.
Local subjects	3, 8, 12	The category posts describe internal news happening within the country and region. The documents belonging to these topics are classified with high probability, which is explained by usage of certain words and phrases related to local sites and their events.
Internal affairs	1, 7, 10	A considerably high number of documents is typical for this category. They represent discussions about internal political and economic issues.
Global geopolitics	0, 6, 13	The topics contain posts with pieces of news and opinions about events happening throughout the world. The main terms include named entities (countries, states, and their leaders).
Other	2, 5	The documents of this category have mainly low probabilities since their subjects of discussion are mixed and can be classified into many small topics.

Despite the efficacy of knowledge extraction and the well-designed dashboard set, this solution faces specific challenges and limitations. Primarily, the Splunk Enterprise is resource-intensive, and the software demands powerful computational capabilities. If these constraints can't be met, open-source visualization and data manipulation tools might be a viable alternative.

The selected LDA algorithm also has inherent drawbacks. It doesn't represent topic evolution over time, necessitating our custom implementation. The algorithm doesn't model sentence structure and disregards sentiment. When analysing opinion data from social media, similar keyword sets might convey opposing sentiments across different documents. Ideally, such observations would be classified as distinct topics. However, the LDA algorithm would likely consider only word usage within documents, resulting in either a single topic or multiple closely related topics. Comprehensive reviews of other topic modelling algorithms validate their successful deployment in various domains [10, 11], and BERTopic emerges as a robust transformer-based neural topic modelling algorithm. The model supports diverse topic modelling activities, including dynamic ones. Another advantage is the integration of GPT models to summarize returned topics [12]. Consequently, topic summarization emerges as a promising research area to glean insights from vast document volumes [13, 14, 15].

4 CONCLUSIONS

This endeavour facilitated real-time monitoring of social media discussions surrounding the Russian war in Ukraine. We ingested and analysed approximately 30GB of data from February 2022 to May 2023 in Splunk using the developed dashboards. We not only modelled a plethora of topics during the study period but also discerned the distinctions among them. The MLOps-specific dashboards for experiment and run evaluations enabled swift determination of the optimal topic count for data segmentation. We postulate that web discussions mirror actual battlefield events and portray the conflict's gravity. Notably, the discovery dashboard and text analysis provide invaluable insights into whether specific text entities and overall topics contain destructive content, which is paramount in the current era of copious opinion data on social media platforms. Another solution advantage is the discovery of top hashtags. On social media, hashtags amplify information's searchability and visibility. We pinpointed frequently used hashtags within specific topics, which is also invaluable for identifying misleading or harmful content. Swift identification of popular text pieces, based on user engagement metrics and their predicted topics, facilitated comprehensive knowledge mining of web discussions about geopolitical events. The MLOps facet of the solution is vital for precise modelling, making it an invaluable tool for media analysis and content moderation.

ACKNOWLEDGMENTS

We want to extend our gratitude to all the people who stand for the freedom of Ukraine and the entire civilized world.

REFERENCES

[1] V. Solopova, O. Popescu, C. Benzmüller, and T. Landgraf, "Automated multilingual detection of pro-kremlin propaganda in newspapers and telegram posts", Datenbank-Spektrum, vol. 23, no. 1, p. 5-14, 2023, doi: 10.1007/s13222-023-00437-2.

[2] M. Popova, E. Siemens, and K. Karpov, "The concept of text processing in an ontological approach to spatio-temporal social network analysis", 2023 30th International Conference on Systems, Signals and Image Processing (IWSSIP), 2023, doi: 10.1109/iwssip58668.2023.10180274.

[3] D. Kreuzberger, N. Kuehl, and S. Hirschl, "Machine learning operations (mlops): overview, definition, and architecture", IEEE Access, vol. 11, p. 31866-31879, 2022, doi: 10.1109/access.2023.3262138.

[4] Studer, S. Bui, T.B. Drescher, C. Hanuschkin, A. Winkler, L., Peters, and et. al, “A Machine Learning Process Model with Quality Assurance Methodology.” Mach. Learn. Knowl. Extr. 2021, 3, 392-413, doi: 10.3390/make3020020.

[5] Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the Space of Topic Coherence Measures. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15). Association for Computing Machinery, New York, NY, USA, 399-408, [Online]. Available: https://doi.org/10.1145/2684822.2685324.

[6] D. Blei, A. Ng, and M. Jordan, “Latent dirichlet allocation”, Journal of Machine Learning Research 3 (2003) 993-1022, doi: 10.5555/944919.944937.

[7] R. Churchill and L. Singh, (2022). The evolution of topic modeling. ACM Computing Surveys, 54(10s),(2022) 1-35, doi: 10.1145/3507900.

[8] S. Fedushko, T. Ustyianovych, and M. Gregus, "Real-time high-load infrastructure transaction status output prediction using operational intelligence and big data technologies", Electronics, vol. 9, no. 4, p. 668, 2020, doi: 10.3390/electronics9040668.

[9] S. Fedushko, T. Ustyianovych, Y. Syerov, and T. Peracek, "User-engagement score and slis/slos/slas measurements correlation of e-business projects

through big data analysis", Applied Sciences, vol. 10, no. 24, p. 9112, 2020, doi: 10.3390/app10249112.

[10] I. Vayansky and S. Kumar, "A review of topic modeling methods", Information Systems, vol. 94, p. 101582, 2020. doi: 10.1016/j.is.2020.101582.

[11] A. Abdelrazek, Y. Eid, E. Gawish, H. Mohamed, and A. Hassan, "Topic modeling algorithms and applications: a survey", Information Systems, vol. 112, p. 102131, 2023, doi: 10.1016/j.is.2022.102131.

[12] "BERTopic". [Online]. Available: https://maartengr.github.io/BERTopic/index.html [Accessed: Jul. 27, 2023].

[13] R. Belwal, S. Rai, and A. Gupta, "Extractive text summarization using clustering-based topic modeling", Soft Computing, vol. 27, no. 7, p. 3965-3982, 2022, doi: 10.1007/s00500-022-07534-6.

[14] A. Alambo et al., "Topic-Centric Unsupervised Multi-Document Summarization of Scientific and News Articles," 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 2020, pp. 591-596, doi: 10.1109/BigData50022.2020.9378403.

[15] A. Joshi, E. Fidalgo, E. Alegre, and L. Fernández-Robles, "Deepsumm: exploiting topic models and sequence to sequence networks for extractive text summarization", Expert Systems With Applications, vol. 211, p. 118442, 2023, doi: 10.1016/j.eswa.2022.118442.

An Exploratory Factor Analysis Approach in Empirical Study of the DigIn.Net 2 Project

Oksana Vasylenko[1], Eduard Siemens[1] and Halyna Henseruk[2]
[1]*Departments of Electrical Engineering, Mechanical Engineering and Engineering Management, Anhalt University of Applied Sciences, Bernburger Str. 55, Köthen, Germany*
[2]*Department of Informatics and Methods of its Teaching, Ternopil Volodymyr Hnatiuk National Pedagogical University, M. Kryvonosa Str. 2, Ternopil, Ukraine*
{*oksana.vasylenko, eduard.siemens*}*@hs-anhalt.de, genseruk@tnpu.edu.ua*

Keywords: Exploratory Factor Analysis, Key Factors of Internationalization, Internationalization of Universities, DigIn.Net Project 2, Surveys.

Abstract: The main purpose of the current DigIn.Net 2 project (supported by DAAD, BMBF) is to promote the digital internationalization of partner universities in Germany and Ukraine. The study focuses on the effective practical application of the project of the Anhalt University of Applied Sciences in cooperation with four Ukrainian universities. Based on the exploratory factor analysis of empirical survey data from 1063 respondents who participated in international internships "Digital Future: Blended Learning" during 2022-2023, an original compilation of key factors for successful internationalization of universities in the digital space was formulated. As the international internship was launched during the pandemic and then continued uninterrupted in the post-pandemic period and during the war, the authors had to collect and analyze valuable data from respondents in difficult times for the first time. In addition, some researchers have been involved in surveys and data processing on internationalization, but they have not applied factor analysis to this type of data. Factor analysis was conducted to reduce the sample size of the variables, group them based on correlations, and interpret the resulting factors. The principal components method was used as the extraction method, and Varimax with Kaiser normalization was used as the rotation method. Thus, a modern model of a successful internationalization strategy for each university should take into account the following factors: Global cooperation and networking, Competences and globalization of standards, Digital evolution, Funding and policy dynamics, Int'l research initiatives.

1 INTRODUCTION

In the dynamic landscape of higher education, the process of internationalization has emerged as a crucial paradigm for universities aiming to foster global competencies and collaborations. With the advent of the digital era, this imperative has been further underscored, propelling institutions to adapt and evolve within an increasingly interconnected and technology-driven milieu.

Aligned with this progressive vision, the DigIn.Net 2 project [1], spearheaded by the German Academic Exchange Service (DAAD), has been instrumental in promoting internationalization endeavors in digital space between partner universities in Germany and Ukraine [2]. This initiative has served as a catalyst for collaborative efforts, particularly exemplified by the Anhalt University of Applied Sciences (HSA) and its joint projects with several prominent Ukrainian universities [3].

This study delves into the critical components and attributes that drive the success of university internationalization in the digital space [4]. Using the method of Exploratory Factor Analysis (EFA), the most probable number and composition of objective factors were obtained, which meaningfully cover the entire complex set of data obtained from participants of the international internships "Digital Future: Blended Learning" internships in 2022 and 2023.

The analytical methodology encompassed the application of the principal components extraction method, complemented by the Varimax technique with Kaiser normalization as the rotation method [5, 6]. This paper highlights a nuanced model for devising and implementing successful

internationalization strategies tailored to the specific needs of each university. The identified factors in the analysis will serve as pivotal cornerstones in shaping an effective and sustainable internationalization trajectory within the contemporary digital landscape. By embracing these multifaceted dimensions, universities can proactively navigate the complexities of globalization and digital transformation, thus fortifying their global presence and impact.

2 INTERNATIONALIZATION OF PARTNER UNIVERSITIES

Research on various dimensions of internationalization is actively conducted worldwide, involving professional organizations, universities, and individual researchers. Notable organizations in this field include the European Association of Universities (EAU), the International Association of Universities (UIA), the British Council (UK), the Institute for International Education (USA), the Center for International Higher Education at Boston College (USA), the Association for Academic Cooperation (USA), the German Academic Exchange Service (Germany), among others.

As per J. Knight's definition, internationalization is the process of incorporating international, intercultural, and global elements into the educational, academic, and administrative functions of an individual institution [7].

Presently, global collaboration stands as a key metric for assessing the caliber of education and scientific pursuits worldwide, serving as a crucial means for their assurance and enhancement. As a result, nearly all higher education institutions globally are actively engaged in international endeavors, striving to broaden their scope, albeit with variations in their interpretation of its principles, aims, objectives, and methodologies.

The German-Ukrainian educational partnership stands out as a promising avenue in this trajectory, particularly in light of the escalating demand for online education and the ongoing large-scale war. This conflict precipitated a significant influx of Ukrainian refugees into Europe, with Germany being a notable destination, rendering the German-Ukrainian partnership increasingly pertinent. As of August 2023, Germany is home to 1 175 695 Ukrainian residents [8], comprising a considerable segment of students and prospective students, whose presence serves as a key metric for measuring the internationalization of universities. In particular, since the beginning of the war, the HSA has hosted more than 220 Ukrainian students, faculty of partner universities, and their families.

2.1 Anhalt University of Applied Sciences

Anhalt University of Applied Sciences (HSA) is a key participant in the dynamic process of internationalization, catering to a diverse cohort of 2 700 students from 110 nations and offering 12 English-language study programs. With international students comprising 35% of the student body, the current internationalization strategy (2018-2024) [9] places emphasis on pivotal areas including fostering innovation, advancing teaching and research, cultivating student competencies, and augmenting external funding.

HSA operationalizes its global focus through five core areas, playing a pivotal role in its continual evolution as a prominent center for education and research. These areas encompass strategic internationalization management, global networking initiatives, the cultivation of an inclusive and welcoming culture, enhancements in learning and teaching, and the promotion of research activities [9].

Expanding participation in various programs and projects serves as a catalyst for supporting innovation, developing research, facilitating the integration of excellent scientists, stimulating student mobility, and significantly increasing external funding.

2.2 HSA-Ukraine Cooperation

HSA maintains robust ties with Ukrainian universities, fostering regular student and academic exchanges. These initiatives are primarily facilitated through programs such as ERASMUS+ (KA 107), the Eastern Partnership, professional international internship programs [10].

The University has 12 English-language programs and has established English-language master's double degree programs such as DDP Communication and Embedded Systems, DDP Data Science, as well as German-language bachelor's programs in Electrical and Information Engineering and Computer Systems and Networks/ Internet of Things (IoT) with the following leading universities:

- National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" (KPI, Kyiv);
- Odessa Polytechnic National University (OPNU, Odesa);

- State University of Intelligent Technologies and Telecommunications (SUITT, Odesa);
- Ternopil Volodymyr Hnatiuk National Pedagogical University (TNPU, Ternopil);
- Kharkiv National University of Radio Electronics (NURE, Kharkiv).

More than 70 students have successfully completed these programs and are currently employed by German and Ukrainian companies. Additionally, more than 40 professors and associate professors have participated in internships at HSA, and over 1100 faculty undertook international online internships in the years 2022-2023.

Through collaborations with Ukrainian universities, especially within the framework of programs administered by the German Academic Exchange Service (DAAD) and funded by the Federal Ministry of Education and Research (BMBF), HSA has accumulated initial and favorable experiences. Various ongoing projects such as the IDEA-East Hub, Digin.Net, Digin.Net 2, DigiJED, DigiJED 2, Study Visits, FIT4Ukraine, GLS Computer Systems and Networks/ Internet of Things (IoT), CaESaR and INTEGRA projects serve to expand and strengthen promising cooperation in the academic field.

2.2.1 DigIn.Net 2 Project

The DigIn.Net Project was initiated shortly before the outbreak of the COVID-19 pandemic, serves as a pivotal testament to HSA's unwavering dedication to international collaboration in higher education and the global orientation of its faculties. It underscores the institution's resolute focus on establishing transnational higher education through the seamless integration of digital technologies and innovations, propelling an accelerated transition towards the future of education.

Building upon the positive momentum garnered during the project's initial phase from 2019 to 2021, the second phase spanning 2021 to 2023 prioritizes the consolidation and expansion of collaborative efforts, accompanied by a steadfast commitment to enhancing the core activities of the project. A notable emphasis is placed on the elevation of digital competencies among the teaching staff.

2.2.2 Project Highlights

Annual *Contests of Innovative Ideas* for students and young scientists in the domains of ICT, energy technologies, and mechanical engineering have been pivotal in showcasing HSA's robust laboratory infrastructure and equipment. This has particularly attracted talented engineering students from Ukraine, presenting them with a platform to actualize their ideas within the laboratory setting, subsequently influencing decisions concerning further study or research endeavors in Germany.

The pivotal highlight of the DigIn.Net 2 project was the introduction of the *"Digital Future: Blended Learning" International Internship* in May 2022, catering to participants from partner universities. A total of 186 participants successfully completed the internship, receiving certificates of completion and enriching their digital competencies through immersive learning experiences. Subsequent iterations of the internship took place in October 2022, April 2023, and October 2023, with a growing number of participants from diverse universities. The internship program served as an enriching platform for comprehending, analyzing, and exploring global facets of internationalization within the digital realm, providing invaluable insights into key factors and perspectives driving this transformative process.

3 EMPIRICAL RESEARCH

The main purpose of this research was to make sure that the fundamental factors that are crucial for the successful internationalization in a digital space of universities are understood. It aimed to explore the significance, challenges, prospects and impact of digital internationalization on learning, as well as on individual actors in the educational process and the nature of inter-institutional relations. Leveraging data from four surveys conducted within the framework of the DigiNet 2 project [2], the study applied exploratory factor analysis to identify key factors. The surveys involved researchers, professors, lecturers, associate professors, and staff of more than 150 higher education institutions.

3.1 Methodology and Respondents

The surveys "Key Factors of Internationalization of Higher Education" were conducted between May 4th and May 28th, 2022, then from October 10th to October 28th, 2022, and then from April 4th to April 22th, 2023, and then from October 2nd to October 20th, 2023. The participants were carefully selected to form a statistically representative sample. Prior to participating in the survey, respondents were provided with an information booklet and a consent form. They were informed that their involvement was entirely voluntary and anonymous, and they had the option to withdraw their consent at any time without

having to provide a reason. It was assured that the collected data would remain confidential and solely be used for research and statistical analysis purposes.

To validate the research tools, a pilot study was conducted on April 4, 2022, involving the participation of 45 teachers and 15 students. Based on the results of the pilot study, adjustments and minor changes were made to the questionnaire items. The decision was made to carry out the survey during online sessions of international internships, utilizing digital tools for data storage, analysis, and presentation.

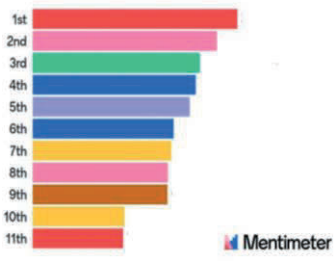
The first survey was conducted from May 4th to May 31st, 2022, and gathered responses from 186 participants representing 15 universities in Ukraine, 3 in Germany, and 5 from other European countries. The second survey took place from October 10th to October 31st, 2022, with 192 respondents from 27 universities. The third survey took place from April 4th to April 22th, 2023, with 261 respondents from more than 100 universities. The fourth survey took place from October 2nd to October 20th, 2023, with 424 respondents from more than 120 universities. The majority of higher education respondents had a public institutional profile, with a dual emphasis on both teaching and research and offering programs at all degree levels.

Questionnaires and a glossary of terms and definitions were made available in both English and Ukrainian. Digital tools, such as Mentimeter, Google Forms, and Slido, were employed for questionnaire creation, design, and result presentation. The findings from the four research rounds are structured into two main segments: an empirical aggregate analysis, comprising the outcomes of the first part of the surveys, and a factor analysis of the empirical data from the second part of the survey.

3.2 Aggregated Results Overview

Aggregated results:

- According to the surveys, an average of 68% of respondents have an institutional policy/strategy, and 8,4% reported that such a policy/strategy is being developed; 23,1% said they had no information about it. A comparison over time indicates a notable rise in the percentage of respondents who reported the implementation of the internationalization policy in a digital space. Specifically, 71,6% of participants reported such a policy in the fourth survey, whereas 64% did so in the first survey. This increase can be interpreted as an objective improvement in the awareness of participants within the educational process.
- Furthermore, the study revealed that the most frequently assessed areas of internationalization are focused on cultivating partnerships, including research collaboration, grant programs, and knowledge sharing/transfer, ranking first. The second most evaluated aspects include the facilitation of faculty mobility, enrollment of foreign students, and promotion of student mobility abroad. Lastly, the development of programs aimed at enhancing national capacity occupies the third position in terms of evaluation priority. Refer to Figure 1 for detailed results.



1st: University partnerships: research collaborations, grant collaborations, program collaborations, knowledge sharing&transfer, patent development, commercialization, faculty and student exchanges

2nd: Cross-border faculty mobility: visiting professorships, faculty exchanges, research project collaborations, expert consultants, and guest lectures

3rd: International training and development programs to build capacity in the countries

4th: Cross-border student mobility: study abroad, student exchanges, student research abroad, and internships abroad

5th: Engagement with/membership in international education organizations (e.g., IIE, UNESCO, NASFA, CIES, IAEA)

6th: Supporting international scholarship opportunities (e.g., Fulbright, DAAD, Erasmus+, Boren, Marshall, Rhodes, Critical Language Scholarships)

7th: Hosting international academic conferences

8th: Domestic-based foreign language preparation/training

9th: Domestic-based international studies programs

10th: Expansion of institutional programs

11th: Domestic-based work with international populations

Figure 1: Survey results: Areas of internationalization.

The results of this survey closely correlate with the data of the latest DAAD study on the areas of internationalization in higher education through the lens of digitalization, where the following are [11]:

1) Cooperation, collaboration and partnerships – CCP.
2) Mobility and Exchange – ME (Physical – Blended – Virtual).
3) Administrating Internationalization Digitally – AID.
4) Attractiveness of Universities in a Global Knowledge Society – GKS.
5) Knowledge Transfer and Open Education – KTOE.

Following the identification of the main areas of internationalization, it was imperative to assess the extent of the respondents' involvement in these areas. The findings revealed the following trends:

- The percentage of respondents engaged in international activities, particularly project-related activities, exhibited a substantial increase from 55,5% in 2022 to 68% in October 2023, and this is most likely to be distance work or blended learning activities;
- In terms of experiences abroad for the purpose of exchange or mobility, the results displayed a slight regression. In October 2022, 49,7% of respondents indicated no experience in this regard, while 38,1% reported internship experiences lasting up to 3 months. Additionally, 7,1% of respondents had experiences spanning from 3 to 6 months, 2,6% had work experiences abroad lasting up to one year, and 2,5% had experiences exceeding one year. In October 2023, these figures changed slightly, with 56% of respondents indicating no experience, 33% reporting internship experiences of up to 3 months, 4% with experiences lasting from 3 to 6 months, 3% having work experiences abroad for up to one year, and 4% having experiences exceeding one year (Figure 2).

3.3 Factor Analysis

Factor analysis compresses information in a correlation matrix, revealing hidden factors that explain relationships among observed features (variables). By identifying a small set of factors, it simplifies complex relationships among numerous variables. These factors group related variables for meaningful interpretation, combining strongly correlated variables while maintaining weak correlations between different factors.

There are two primary classifications of factor analysis: Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) [12]. EFA is particularly beneficial when the researcher lacks a clear understanding of the specific number of factors at play [13].

The mathematical model of factor analysis assumes changes in observed variables stem from underlying latent properties known as ***general factors***, with each common factor having varying degrees of influence denoted by ***factor loadings***. Additionally, it acknowledges that variations can also arise from independent sources, termed ***specific factors***, such as random errors and measurement inaccuracies. Let X_i represent the i-th observed variable. We can denote U_i as the general component of this variable, which comprises the changes influenced by common factors. Additionally, ε_i represents the specific factors, which entail changes that are independent of one another and not contingent on alterations in other variables. Consequently, the change in variable X_i can be dissected into the sum of the general component, representing overall changes influenced by common factors, and the specific component, indicating changes induced by specific factors:

$$X_i = U_i + \varepsilon_i \tag{1}$$

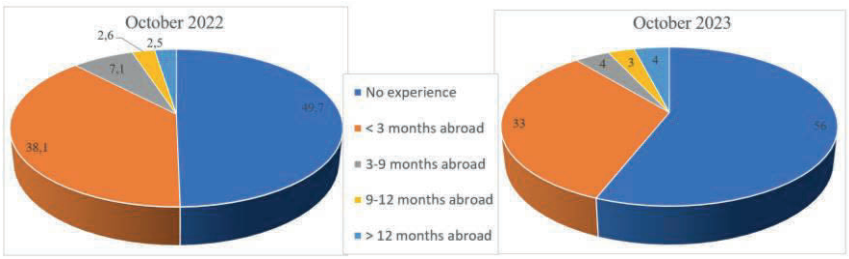


Figure 2: Survey results: International experience and mobility of the respondents in 2022 and 2023.

The further improvement of the idea of factor analysis is based on the assumption that the data of n variables U_i are linear combinations of a smaller number of other variables F_j called factors, i.e.

$$U_i = \omega_{1i}F_1 + \omega_{2i}F_2 + ... + \omega_{ki}F_k, \quad (2)$$

where $i = \overline{1,n}$.

ω_{ji} - are factor loadings of the F_j factors, which characterize the degree of influence of the j-th common factor on the i-th empirical variable. Combining the above (1) and (2), we obtain the following, which is expressed through general and specific factors as follows:

$$X_i = \omega_{1i}F_1 + \omega_{2i}F_2 + ... + \omega_{ki}F_k + \varepsilon_i, \quad (3)$$

where $i = \overline{1,n}$.

Where:

- General factors F_j are either uncorrelated random variables with variance equal to 1 or unknown non-random parameters.
- The specific factors ε_i have a normal distribution, are uncorrelated and independent of the general factors.

The primary method for directly estimating the relationship between variables (3) typically assumes equal relationship scores of 1. In this approach, the relationship scores are equivalent to the diagonal elements of the correlation matrix, resulting in specific factor variances of 0. This concept forms the foundation of one of the most renowned models of factor analysis, known as the principal components method [5]. Under this method, a transition to a new coordinate system occurs, constituting a set of orthonormalized linear combinations. These linear combinations are specifically the eigenvectors of a correlation matrix. The first principal component is the linear combination exhibiting the highest variance. Subsequently, the second component embodies the largest variance among all other linear combinations that exhibit no correlation with the first principal component.

The next stage of factor analysis is factor rotation, i.e. rotation of the coordinate axes corresponding to the factors, which is carried out not in the space of the original variables, but in the space of the factors found. The essence of the process of rotation of any pair of vectors is to find an angle between the new and old direction of the factors that would give the greatest increase in the selected criterion. Rotation of factors in space allows to characterize each attribute by the predominant influence of one factor. In modern statistical data processing packages, the most common rotation methods are varimax, quartimax, and equimax. Varimax rotation simplifies the values of the columns of the factor matrix by reducing them to 1 or 0. Having obtained the factor solution (factor matrix) after the rotation procedure, you can proceed to the interpretation and naming of the factors.

3.3.1 EFA Methodology

Factor analysis was conducted to reduce the size of the sample of features (answers to the questionnaire) and group them based on correlations. The principal components method was used as a selection method, and Varimax with Kaiser normalization was used as a rotation method [6]. To optimize the calculations, the statistical software package SPSS was used [14], the results of which are presented in the following Tables 1 and 2.

The survey was conducted in four rounds, which were combined into two blocks and compared over time: in 2022 with 378 respondents and in 2023 with 685 respondents, respectively. Respondents were asked to express their opinion on the following items:

1) Areas of digital internationalization development:
   - *AD1*. Academic partnerships and collaboration;
   - *AD2*. Mobility and exchange programs;
   - *AD3*. Digital transformation;
   - *AD4*. Global standards implementation;
   - *AD5*. International research initiatives.
2) External factors of internationalization development:
   - *EF1*. Public policy (national, state/provincial, municipal);
   - *EF2*. Ranking of the country's universities;
   - *EF3*. Business and industry;
   - *EF4*. International rankings;
   - *EF 5*. International partner universities.
3) Internal factors of internationalization development:
   - *IF1*. Head of the institution;
   - *IF2*. International *department or person responsible for internationalization;*
   - *IF3*. Teaching staff;
   - *IF4*. University ranking;
   - *IF5*. Personally.
4) Internal and external challenges in the development of internationalization:
   - *CH1*. Insufficient financial resources;
   - *CH2*. Restricted government funding;

- *CH3*. Limited awareness among educational stakeholders;
- *CH4*. Faculty and staff's limited experience and knowledge;
- *CH5*. Language barrier.

5) Challenges and solutions of the digital internationalization:
   - *CD1*. Overcoming the consequences of the COVID-19 pandemic and global conflicts;
   - *CD2*. Rethinking education at all stages of learning;
   - *CD3*. Implementation of hybrid learning models (face-to-face, blended, virtual);
   - *CD4*. Promoting e-internationalization and mixed virtual mobility;
   - *CD5*. Promoting digital collaboration and virtual partnerships.
6) Implementation and expansion of external activities:
   - *EA1*. Cultivate international networks;
   - *EA2*. Cultivate intercultural and global competencies through knowledge exchange;
   - *EA3*. Facilitate joint research endeavors;
   - *EA4*. Collaborate with international associations and institutions;
   - *EA5*. Establish international or double degree programs;
   - *EA6*. Develop cross-border academic programs;
   - *EA7*. Engage in joint projects.
7) Preparation and bridging gaps in the path of digitalization:
   - *GD1*. Creating an innovative digital environment;
   - *GD2*. Digitization of educational and research processes;
   - *GD3*. Enhancing digital competencies;
   - *GD4*. Acquiring practical experience in digital technologies;
   - *GD5*. Augmenting the amount of external funding.

The respondents have ranked the proposed answers in descending order of importance, i.e. from the complete preference of one answer over another. After that, each of the proposed answers to each question was evaluated on a five-point (seven-point) scale: from strongly agree (2/3) to strongly disagree (-2/-3).

3.3.2 Implementation

Based on the results of the factor analysis, Table 2 indicates the total explained variance, revealing that five Eigen factors have values exceeding one, leading to the selection of only five factors for analysis. The first factor explains 27,113% of the total variance, the second one explains 7,174%, the third explains 7,033%, the fourth one explains 6,121%, and the fifth – 4,424%. Following the extraction phase, the Kaiser's method, also known as K1 - Kaiser's method [6], was employed to determine the number of constructs to retain for rotation. As shown in Table 1, the analysis revealed that these five factors should be retained for further study.

Table 1: Total explained variance.

	Initial Eigenvalues			Extraction Sums of Squared Loading		
	Total	% of Variance	Cumula-tive % of Var.	Total	% of Variance	Cumula-tive % of Var.
1	6,158	27,113	27,113	6,158	27,113	27,113
2	1,629	7,174	34,287	1,629	7,174	34,287
3	1,597	7,033	41,32	1,597	7,033	41,32
4	1,390	6,121	47,441	1,390	6,121	47,441
5	1,005	4,424	51,865	1,005	4,424	51,865
6	0,934	4,113	55,978			
7	0,913	4,021	59,999			
8	0,814	3,584	63,583			
9	0,692	3,045	66,628			
10	0,549	2,417	69,045			
11	0,525	2,311	71,356			
12	0,500	2,202	73,558			
13	0,497	2,187	75,745			
…	…	…	…			
36	0,073	0,321	99,802			
37	0,045	0,198	100			
Extraction Method: Principal Component Analysis						

In the next step, the factor loadings with the highest absolute values were identified and recorded in each row of the rotated component matrix (Table 2). These factor loadings should be interpreted as correlation coefficients between the variable and the factors. For example, variable *AD1* has the strongest correlation with factor 1 (correlation coefficient 0,616), variable *AD2* has the strongest correlation with factor 1 (0,543), while variable *AD3* is mainly related to factor 2 (0,828), etc. In most

cases, the assignment of a variable to a particular factor based on correlation coefficients is unambiguous.

Table 2: Rotated component matrix.

	Components				
	1	2	3	4	5
AD1	0,616	0,208	0,066	0,323	−0,191
AD2	0,543	0,004	0,231	0,127	0,044
AD3	0,111	0,828	0,205	−3,8E-02	0,067
AD4	0,004	0,031	0,735	0,302	0,167
AD5	0,031	−0,097	0,333	0,602	0,042
EF1	0,200	0,118	0,015	0,123	0,622
EF2	0,279	0,007	−4,2E-02	0, 322	0,054
EF3	0,083	0,022	0,103	0,118	0,761
EF4	0,172	0,066	0,557	0,082	0,244
EF5	0,602	−0,033	0,228	0,174	0,321
IF1	0,592	0,006	0,301	0,202	0,043
IF2	0,870	0,045	0,098	0,124	0,322
IF3	0,233	0,082	0,612	0,202	0,337
IF4	0,613	0,233	0,095	0,118	0,089
IF5	−3,2E-02	0,063	0,204	0,704	0,103
CH1	0,201	0,442	0,205	0,082	0,777
CH2	0,099	0,034	−2,3E-02	0,176	0,811
CH3	0,102	0,087	0,605	0,320	0,166
CH4	0,089	0,201	0,553	0,118	0,088
CH5	0,222	0,096	0,751	0,044	0,117
CD1	0,081	0,238	0,113	−0,282	0,644
CD2	0,054	0,007	0,665	0,062	0,224
CD3	0,073	0,911	−0,202	0,163	0,314
CD4	0,300	0,628	0,105	0,062	0,105
CD5	−0,104	0,823	0,114	−0,112	0,088
EA1	0,733	0,211	0,073	0,332	0, 094
EA2	0,211	0,074	0,704	0,134	−0,072
EA3	0,210	0,081	−0,005	0,814	0,088
EA4	0,769	0,122	0,204	0,311	0,099
EA5	0,821	0,008	0,213	0,092	0,104
EA6	0,014	0,172	0,666	0,200	0,116
EA7	0,092	0,019	0,301	0,812	0,099
GD1	0,222	0,628	0,104	0,088	−0,207
GD2	0,172	0, 691	0,076	−0,107	0,205
GD3	0,098	0,028	0,752	0,189	0,169
GD4	0,186	0,567	0,123	0,302	0,088
GD5	0,077	0,028	0,201	−0,201	0,607
Extraction Method: Principal Component Analysis Rotation Method: Varimax with Kaiser Normalization a. Rotation converged in 8 iterations					

3.3.3 Results Interpretation

Based on the results, one variable (*EF2*) should be eliminated as its loadings are below 0,5.

The factor matrix data (rotated component matrix) indicate 5 key factors for successful internationalization of the university in an innovative digital environment. To interpret each factor, the analysis of the variables included in it should be conducted to identify common characteristics. The more variables with high loadings in a factor, the easier it is to discern its nature. While there are no standardized methods for selecting factor names, a preliminary approach could involve using the name of the variable with the highest loadings within the factor.

Therefore, based on the above, the variables can be attributed in the following order to the factors:

- ***Factor 1 – Global Cooperation and Network (GCN), 27,03%*** consists of: *AD1, AD2, EF 5, IF1, IF2, IF4, EA1, EA4, EA5.*
- ***Factor 2 (18,92%) – Digital Evolution (DEv)*** includes variables *AD3, CD3, CD4, CD5, GD1, GD2, GD4.*
- ***Factor 3 (24,32%) – Competencies and Globalization of Standards (CGS)*** includes variables *GD3, CD2, EA2, EA6, AD4, IF3, CH3, CH4, CH5.*
- ***Factor 4 (10,81%) – Int'l Research Initiatives (IRI)*** includes variables *AD5, EA3, EA7, IF5.*
- ***Factor 5 (16,22%) – Funding and Policy Dynamics (FPD)*** includes variables *EF1, EF3, CH1, CH2, CD1, GD5.*

Thus, it is possible to outline and demonstrate a model of key factors for successful internationalization of a university in an innovative digital environment from the perspective of active participants in the educational process (Figure 3).

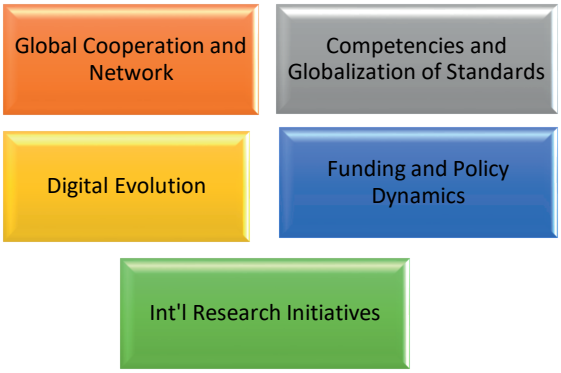


Figure 3: Key factors of university internationalization in the digital space.

4 CONCLUSIONS

The study, which focused on the practical implementation of the DigIn.Net 2 project of the HSA in cooperation with four Ukrainian universities, conducted an exploratory factor analysis of empirical

survey data from 1063 respondents who participated in the international internships "Digital Future: Blended Learning" in 2022-2023. This study faced the unprecedented challenge of collecting and analyzing data during the pandemic and wartime conditions, highlighting the resilience of the research process. Additionally, the application of factor analysis to the dataset, a novel approach in this context, enabled the researchers to condense variables, identify correlations, and interpret key factors efficiently.

As a result, a comprehensive set of key factors for the successful internationalization of universities in the digital sphere was identified. These factors include the need to actively develop (1) global cooperation and network, (2) competencies and globalization of standards, (3) digital evolution, (4) funding and policy dynamics, and (5) international research initiatives. These insights offer an innovative framework that can guide the development and implementation of effective internationalization strategies for universities, emphasizing the significance of adapting to the evolving digital landscape while also fostering global collaboration and maintaining a strong emphasis on research initiatives. By acknowledging these key factors, institutions can fortify their position in the international academic community, nurturing a dynamic environment that promotes growth, excellence, and sustainable global engagement.

ACKNOWLEDGMENTS

This paper was prepared at the Anhalt University of Applied Sciences with the support of the DigIn.Net 2 project - 57602278, which is managed by the German Academic Exchange Service (DAAD) and funded by the Federal Ministry of Education and Research (BMBF).

REFERENCES

[1] DigIn.Net 2, “German-Ukrainian Digital Innovation Network Project 2,” 2023, [Online]. Available: https://digin-net.de/en/.

[2] DAAD, “Unterstützung der Internationalisierung ukrainischer Hochschulen, ” 2023, [Online]. Available: https://www.daad.de/de/infos-services-fuer-hochschulen/weiterfuehrende-infos-zu-daad-foerderprogrammen/internationalisierung-ukrainischer-hochschulen/.

[3] Anhalt University of Applied Sciences, “International partner universities of Anhalt University, ” 2023, [Online]. Available: https://www.hs-anhalt.de/en/international/studyexchange-and-internship-abroad/partner-universities.html.

[4] S. Chang and C. Gomes, “Why the Digitalization of International Education Matters,” Journal of Studies in International Education 26 (2), 2022, pp. 119-127, [Online]. Available: https://doi.org/10.1177/10283153221095163.

[5] J. Bai and S. Ng, “Principal components estimation and identification of static factors,” Journal of Econometrics Volume 176, Issue 1, September 2013, pp. 18-29, [Online]. Available: https://doi.org/10.1016/j.jeconom.2013.03.007.

[6] H. Kaiser, “The application of electronic computers to factor analysis,” Educational and psychological measurement. 1960 Apr; 20 (1): 141-51.

[7] J. Knight, “Higher Education in Turmoil: The Changing World of Internationalization,” Toronto, 2008, 254 p.

[8] Eurostat, “Temporary protection for persons fleeing Ukraine - monthly statistics, ” EU, October 2023. [Online]. Available: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Temporary_protection_for_persons_fleeing_Ukraine_-_monthly_statistics.

[9] Anhalt University of Applied Sciences, “Internationalisierungsstrategie 2018 – 2024, ” Mai 28, 2018, [Online]. Available: https://www.hs-anhalt.de/fileadmin/Dateien/Praesidium/Internationalisierungsstrategie_Formatierung_210518.pdf.

[10] C. Scott, O. Vasylenko, “Mathematical and Statistical Methods of Analyzing the Successful Implementation of German-Ukrainian Projects,” Proceedings of the 11th International Conference on Applied Innovations in IT. 2023, pp. 151-160.

[11] DAAD, “Moving target digitalisation 2020: re-thinking global exchange in higher education,” Oct 5-6, 2020, [Online]. Available: https://www.daad.de/en/the-daad/what-we-do/moving-target-2020/.

[12] H. Taherdoost and M. Madanchian, “Empirical Modeling of Customer Satisfaction for E-Services in Cross-Border E-Commerce, ” Electronics, 2021 Jan; 10 (13): 1547.

[13] H. Taherdoost, “Exploratory Factor Analysis; Concepts and Theory. Advances in Applied and Pure Mathematics“, 2014, pp. 375-382.

[14] O. Vasylenko and I. Sencha, “Mathematical and statistical methods of analysis in applied research,” Textbook. Odesa: ONAT, 2011, 166 p, [Online]. Available: http://ir.nuozu.edu.ua:8080/jspui/handle/lib/319.

Prediction of Child Birth Delivery Mode Using Hybrid-Boosting Ensemble Machine Learning Model

Mamatha Sekireddy and Sudha Thatimakula
Research Scholar, Department of Computer Science, Sri Padmavati Mahila Visvavidyalayam, Padmavathi Nagar, Tirupati, Andhra Pradesh, India
mamatha.sekireddy@gmail.com, *thatimakula_sudha@yahoo.com*

Keywords: Ensemble Methods, Machine Learning, Artificial Intelligence, Voting.

Abstract: Maternal health is a critical aspect of public health that affects the well-being of both mother and infant. Despite the medical breakthroughs, maternal mortality rates remain high, particularly in developing countries. AI-based models provide new ways to analyze and interpret medical data, which can ultimately improve maternal and fetal health outcomes. The present study proposes a hybrid model for the maternal mode of delivery classification in pregnancy, which utilizes the strength of Ada Boost, Gradient Boost, and XG Boost algorithms. The proposed model culminates the three algorithms to improve the accuracy and efficiency of childbirth delivery classification method in pregnant women. The dataset used in this study consists of features such as age, systolic and diastolic blood pressure, blood sugar, body temperature, heart rate, placenta previa, and gravida. The dataset is divided into training and testing sets, where 70% of the data is used for training and the rest 30% for testing. The output of the Ada Boost, Gradient Boost, and XG Boost classifier is considered, and a maximum probability voting system selects the output with the highest probability as the most correct one. Performance is evaluated using various metrics, such as accuracy, precision, recall, and F1 score as an outcome where the XG Boost showed an accuracy of 97.07%, the Ada Boost showed 85.02%, and the Gradient Boost showed 92.51% respectively. Results also conspicuously showed that the proposed model achieves the highest accuracy of t98.05%, with 97.9% precision, 97.9% recall, and an F1 score of 97.9% on the testing dataset. The hybrid model proposed in this study has the potentiality to improve the accuracy and efficiency of the maternal mode of delivery classification in pregnancy, leading to better health outcomes for pregnant women and their babies.

1 INTRODUCTION

In recent years, artificial intelligence (AI) has emerged as a powerful tool in health care, offering new ways to analyze and interpret the complex medical data. AI-based models have shown promising results in various clinical applications, including disease diagnosis, treatment planning, and patient monitoring. AI-based modelscan analyze vast amounts of health data, both structured and unstructured, to identify patientswho may be at high risk for adverse outcomes. These models can help health care providers makemore accurate and timely decisions; and when it refers to maternal health care in pregnancy; it can ultimately improve maternal and fetal health outcomes. AI offers novel approaches to prediction modelling, diagnosis as well as early detection of mode of delivery in pregnancy.

In general, there are two main types of baby delivery: vaginal delivery, often known as normal delivery and caesarean delivery. Each delivery has advantages and disadvantages. Vaginal delivery is the most common method of childbirth because it is associated with little risk. Pregnant women should take care of their health after caesarean deliveries because they can result in blood loss, infections, and difficulties with subsequent pregnancies [1].

A machine learning-based maternal health care system will anticipate the baby's mode of birth in advance through the analysis of the pregnant woman's medical features generated during the pregnancy period.

2 LITERATURE SURVEY

This section briefly summarizes prior research on the prediction of baby delivery. The quadratic discriminant analysis approach demonstrated the maximum accuracy of 0.979992 with the F1 score of 0.979962 in the study by the author [2] on predicting the right mode of birthing using a machine learning algorithm. The authors in [3] proposed a paper on a comparative analysis of supervised machine learning techniques for diagnosing mode of delivery in the medical sciences, in which the Random Forest algorithm gets the greatest accuracy of 99%. [4] Syed Ali Abbas proposed a study on caesarean analysis and the application of Machine Learning methods for birth data classification. The primary causes of caesarean delivery are of being determined to be old age, high blood pressure, and diabetes. A study on the comparative analysis of Naive Bayes and Decision Tree C4.5 for CS prediction was proposed by Gusti Ayu Suciningsih. Memory utilisation, programme execution time, and accuracy metrics are employed in a comparison analysis of the Nave Bayes & Decision Tree given in [5].

3 METHODOLOGY

2.1 Data Description and Operating Environment

The maternal mode of delivery dataset utilized in this study is acquired from the maternal health care center. The dataset contained eight predictive variables in pregnancy viz., Age (in years), Systolic Blood Pressure (in mmHg), Diastolic Blood Pressure (in mmHg), BloodSugar (in mmol/L), Body Temperature (in degrees Fahrenheit), and Heart Rate (in beats per minute), Gravida (in 1,2,3,4) and placenta previa (in yes/no) all of which were numerical variables except placenta previa. A categorical predict or variable, Placenta previa, is present in two possible cases Yes and No. A categorical target variable, Delivery Method, is present in two possible classes; Caesarean, and Normal. The second sequent totaled 9 variables in the whole dataset. There were 1,021 instances in the dataset; 529 samples of which were in the Caesarean class, and 492 samples are in the normal class, as shown in Figure 1. Attributes and their data types in the dataset are specified in Table 1. There are no missing values in the dataset. All samples with all of their features are used in the model building and evaluation. All data manipulations, model building and evaluations are performed in the python programming language (version 3.8.3). Packages utilized include pycaret (2.3.10), numpy (1.20.3), pandas and (1.5.2), sklearn (0.23.2).

Table 1: Dataset specification.

S.no	Attributes	Datatype
1	Age (Input)	Numerical
2	Systolic Blood Pressure (Input)	Numerical
3	Diastolic Blood Pressure (Input)	Numerical
4	Blood Sugar (Input)	Numerical
5	Body Temperature (Input)	Numerical
6	Heart rate (Input)	Numerical
7	Gravida (Input)	Numerical
8	Placenta Previa (Input)	Categorical
9	Delivery Method (Output)	Categorical

Further feature statistics are shown in Table 2 that are based on count, mean, standard deviation, min (minimum values), 25%, 50%, 75%, and max(maximum values). The minimum value shows the lowest limit and the maximum value shows the highest limit for all features.

Table 2: Description of statistical distribution of Dataset Features.

Attribute	Count	Mean	Std	Min	25%	50%	75%	Max
Age	1021	28.6	8.7	18	21	28	35	45
Systolic BP	1021	114.8	20.4	70	95	120	130	160
Diastloic BP	1021	78.6	15.2	49	65	80	90	100
Blood Sugar	1021	9.8	3.9	6	7	7.7	11	19
Body Tempe-rature	1021	98.6	1.4	98	98	98	98	103
Heart rate	1021	74.6	8.0	60	76	76	80	90
Gravida	1021	1.6	0.7	1	2	2	2	4
Placenta Previa	1021	0.1	0.3	0	0	0	0	1
Delivery Method	1021	0.5	0.4	0	1	1	1	1

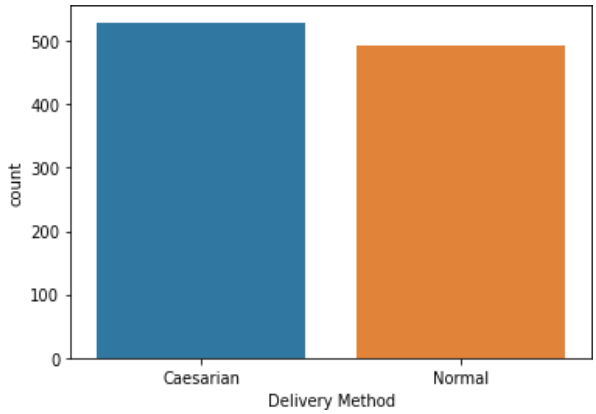


Figure 1: Distribution of dataset based on Delivery Method.

3.2 Data Preprocessing and Analysis

The data file is saved in the comma-separated value (CSV) format, and this facilitated easy reading into the Python script. The major pre-processing step carried out is in the way of encoding the target (categorical) variables and predictor variable placenta previa into numerical variables to facilitate computation. Placenta previa categorical variables yes and no are coded as 1 and 0, and Normal and Caesarean classes were coded as 0 and 1 respectively. Subsequently, the correlation [6] between variables is examined, as depicted in Figure 2. The variable with the highest correlation to the delivery method is found to be blood sugar, while the variable with the lowest correlation is body temperature and heart rate.

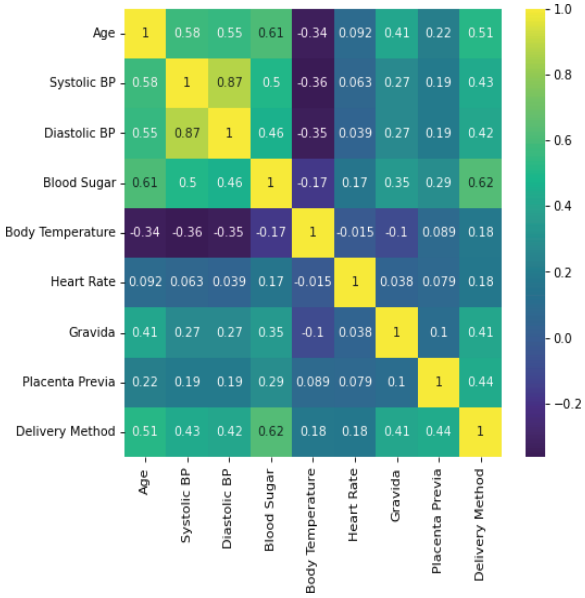


Figure 2: Correlation between the Dataset Variables.

3.3 Machine Learning Ensemble Methods

The data was randomly divided into training and testing sets, with 70% and 30% of the data, respectively. To facilitate data analysis, the Standard Scaler is applied to scale the data. The predictive abilities of various commonly used machine learning ensemble algorithms are initially examined-including, Gradient Boosting Classifier, Ada Boost Classifier, and XG Boost Classifier [7]. A comparison is made among these different ML algorithms using the accuracy, precision, recall, and F1score. The results are summarized in Table 3. The algorithm that performed best according to these metrics is the XG Boost Classifier.

3.4 Training of the Hybrid Model

Afterward, these three classifiers are merged into one final hybrid model. The hybrid model utilized a maximum probability voting technique, this is by looking at the predictions (classes and probability values) of Ada boost, Gradient Boosting, and the XG Boost classifier, it subsequently selects and outputs whichever class had the highest probability value regardless of the difference in class predictions of both classifiers. The preference for this probabilistic approach stems from its demonstrated ability to offer simplicity and good interpretability, as similar maximum likelihood ensembles have shown in the past [8]. Additionally, since only 3 classifier units are employed in this binary-class prediction ensemble, the maximum probability voting techniquecapitalizes on their distinct natures, to the advantage that they will not commit the same errors [9]. Thus, the result from the more confident classifier is selected as the most correct one.

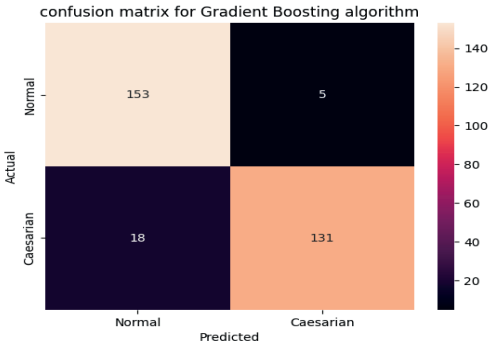
4 RESULTS

The performance of the proposed hybrid model has been evaluated based on the following key metrics: Accuracy, Recall, Precision, and F1 Score [10]. An analysis of the detailed comparison of metrics between Ada boost, Gradient Boost, XG Boost and Hybrid ensemble algorithm provided in Table 3 which reveals that the overall hybrid model performed significantly better with an accuracy score of 0.98. This increase in accuracy could be attributed to the fact that both models in the ensemble differ in the nature of their misclassifications, consequently drawing on each's distinctive nature; to the advantage of one another, to provide the most correct

classification. The precision, recall and F1-score from our hybrid model ranged to 0.9700.The confusion matrix, illustrated in Figure 6 that provides valuable insights into hybrid model performance regarding classification of instances into normal and cesarean delivery methods. The analysis reveals that the model accurately predicted 155 instances as Normal, 146 instances as Caesarean. Of particular significance is hybrid algorithm has an ability to avoid misclassifying any Normal instance as Caesarean. This outcome is crucial, especially in real-world medical applications, as misclassifying a Normal instance as Cesarean could have severe consequences. The model's ability to avoid such errors indicates its reliability and suitability for medical decision-making.

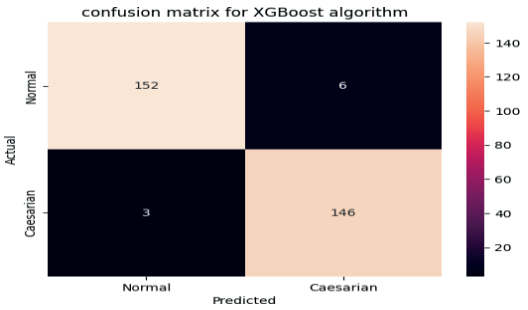
Table 3: Performance comparison of Ada Boost, XG Boost, Gradient Boosting, and Hybrid ensemble model.

Model	Classes	Precision	Recall	F1 Score	Accu-racy
AdaBoost	Normal	0.90	0.80	0.85	0.85
	Cesarean	0.81	0.91	0.85	
Gradient Boost	Normal	0.89	0.97	0.93	0.93
	Cesarean	0.96	0.88	0.92	
XG Boost	Normal	0.98	0.96	0.97	0.97
	Cesarean	0.96	0.98	0.97	
Hybrid boosting ensemble model	Normal	0.98	0.98	0.98	0.98
	Cesarean	0.98	0.98	0.98	



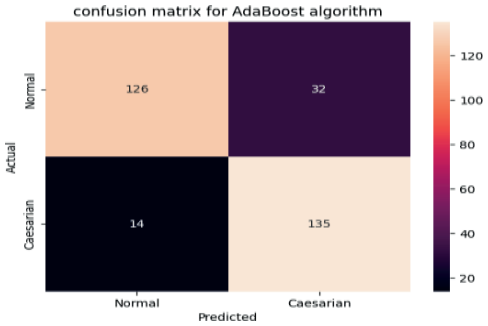
	Precision	Recall	F1-Score	Support
0	0.89	0.97	0.93	158
1	0.96	0.88	0.92	149
accuracy			0.93	307
macro avg	0.93	0.92	0.92	307
weighted avg	0.93	0.93	0.92	307

Figure 3: Gradient Boosting: confusion matrix and classification report.



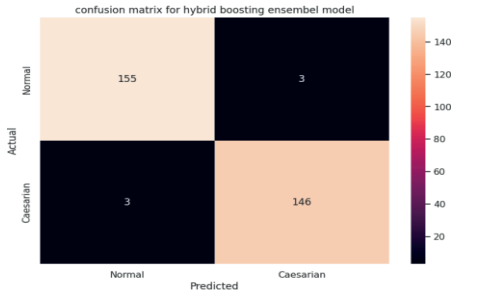
	Precision	Recall	F1-Score	Support
0	0.89	0.97	0.93	158
1	0.96	0.88	0.92	149
accuracy			0.93	307
macro avg	0.93	0.92	0.92	307
weighted avg	0.93	0.93	0.92	307

Figure 4: XG Boost: confusion matrix and classification report.



	Precision	Recall	F1-Score	Support
0	0.90	0.80	0.85	158
1	0.81	0.91	0.85	149
accuracy			0.85	307
macro avg	0.85	0.85	0.85	307
weighted avg	0.86	0.85	0.85	307

Figure 5: Ada Boost : confusion matrix and classification report.



	Precision	Recall	F1-Score	Support
0	0.98	0.98	0.98	158
1	0.98	0.98	0.98	149
accuracy			0.98	307
macro avg	0.98	0.98	0.98	307
weighted avg	0.98	0.98	0.98	307

Figure 6: Hybrid Boosting ensemble model: confusion matrix and classification report.

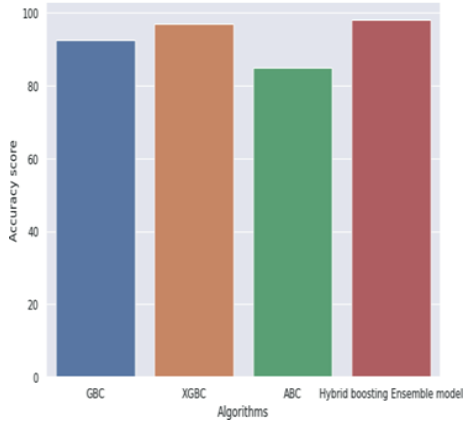


Figure 7: Comparison of Proposed Hybrid model with Ensemble Machine Learning methods.

4 CONCLUSIONS

In conclusion, the present study developed a novel hybrid ensemble model for the maternal mode of baby delivery classification that achieved high accuracy and performed comparably to the previous models. The model has important implications for clinical practice, and the findings suggest that the ensemble technique of machine learning models can be the useful tools for maternal baby delivery classification. Future research is needed to confirm the generalizability of the present findings and to optimize the use of structured and unstructured data in these models. By accurately identifying Cesarean delivery in advance, health care providers can take preventative measures to reduce the likelihood of complications and improve maternal and fetal outcomes. The high accuracy of the model suggests that an ensemble of machine learning models has the potentiality to be the useful tools in the maternal mode of baby delivery classification.

REFERENCES

[1] M. Najrul Islam, "Exploring Machine learning algorithms to find the best features for predicting mode of child birth," Digital Object Identifier 10.1109/ACCESS.2020.3045469.

[2] Md. Kowsher, "Predicting the Appropriate Mode of Childbirth using Machine Learning Algorithms," IJACSA, vol. 12, no. 5, 2021.

[3] S. S. Hussain, "A Comparative Study of Supervised Machine Learning Techniques for Diagnosing Mode of Delivery in Medical Sciences," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 10, no. 12, 2019.

[4] S. A. Abbas, "Cause Analysis of Caesarean Sections and Applications of Machine Learning Methods for Classification of Birth Data," Digital Object Identifier 10.1109/ACCESS. 2018.2879115, IEEE Access, 2018.

[5] G. A. Suciningsih, "Comparative analysis of Naïve Bayes and Decision Tree C4.5 for Cesarean Section Prediction," Journal of Soft Computing Exploration, ISSN: 2746-0991.

[6] A. Crossman, "Correlation Analysis in Research," from www.thoughtco.com: https://www.thoughtco.com/what-is-correlation-analysis-3026696.

[7] Rocca, "Ensemble-methods-bagging-boosting-and-stacking," from towardsdatascience.com: https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205.

[8] K. Dembczynski, W. Kotłowski, and S. Słowi ´ nski, "Maximum Likelihood Rule Ensembles," in Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland: Association for Computing Machinery, p. 224–31. doi: 10.1145/1390156.1390185.

[9] F. Leon, S. A. Floria, and C. Ba˘dica˘, "Evaluating the Effect of Voting Methods on Ensemble-based Classification," in 2017 IEEE International Conference on Innovations in Intelligent Systems and Applications (INISTA), Gdynia: Institute of Electrical and Electronics Engineers (IEEE). doi: 10.1109/INISTA.2017.8001122.

[10] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," arXiv.

A Data-Driven-Based Wide-Area Protection Scheme for Fault Detection Using the Limited Measurements

Sirwan Shazdeh, Qobad Shafiee and Hassan Bevrani
Smart/Micro Grids Research Center, University of Kurdistan, Pasdaran Boulevard, PO Box 416, Sanandaj, Iran
s.shazdeh@uok.ac.ir, q.shafiee@uok.ac.ir, h.bevrani@uok.ac.ir

Keywords: Fault Detection, Wide-Area Protection, Data-Driven Scheme, Intelligent Electronic Devices (IEDs), Microgrid (MG).

Abstract: This paper presents a novel and efficient approach for wide-area fault detection in microgrids, utilizing data-driven techniques based on voltage and current measurements. The proposed method offers both high speed and accuracy in detecting faults. The methodology consists of three key steps that collectively form a comprehensive protection scheme. Initially, the current trajectories obtained from the measurements are analyzed to determine the fault condition. This initial indicator serves as a valuable starting point for fault detection. In the second step, the impedance of the lines, including the considered area, is calculated for the fault detection. The change of the calculated impedances implies for the fault occurrence activating the third step. In the final step, an iterative process is followed to identify the faulted line. The proposed method provides a faster and more reliable fault detection mechanism, allowing for rapid response and mitigation of potential disruptions. The efficacy of the proposed method is validated on an 11-bus microgrid. The simulation investigations are conducted in MATLAB\SIMULINK environment.

1 INTRODUCTION

Due to the escalating concerns of global warming, economic and power loss issues, and the growing demand for electricity, the utilization of renewable energy sources (RESs) as micro-sources in close proximity to the load has witnessed a significant surge at the distribution level. The integration of electrical loads and distributed generations (DGs) forms microgrids (MGs) that operate in both grid-connected and islanded modes [1]. However, despite the evident advantages of MGs, they face a multitude of technical problems and challenges, particularly in the areas of control, operation, and protection [2], [3]. Notably, this paper focuses on the protection challenges within MGs.

The reliable and secure operation of protection systems in MGs is severely affected by various factors, including different short circuit levels in grid-connected and islanded modes, limited current capacity of inverter-based DGs due to insulation issues, the intermittent nature of RESs, and continuous changes in the grid topology [4]. These factors collectively undermine the efficacy of protection systems, and any unplanned disconnection of equipment in MGs, resulting from inadequate performance of protection systems, can lead to unstable conditions due to the limited capacity of DGs.

Consequently, extensive research efforts have been devoted to addressing these challenges [5, 6]. Various adaptive-based protection approaches, ranging from decentralized to centralized actions, have been introduced to determine suitable settings for protection functions [5]. Furthermore, significant contributions have been made to enhance vulnerable protection systems such as distance, overcurrent, and differential functions. Notable advancements in this area can be found in prior works, which build upon the conventional principles of protection devices [7 - 11]. However, challenges persist, including the need for information from the opposite end of the line, communication vulnerabilities, the complexity of setting determination in the face of frequent topology changes, and issues related to the reliability and functionality of protection devices.

In addition to conventional protection-based approaches, several signal-processing-based schemes have been employed to extract relevant features for

accurate fault detection [12]. However, the correct performance of signal-processing-based approaches is hindered by challenges such as intricate threshold determination, susceptibility to measurement errors, and the complex operational conditions of MGs.

Given the significant advancements in artificial intelligence (AI) techniques, AI-based approaches have garnered increased attention for fault detection and identification, particularly in light of the sophisticated and unpredictable conditions encountered in MG operations [13]. For instance, a fault detection method based on support vector machines has been proposed in [14]. Furthermore, the combination of AI-based approaches with knowledge-based schemes has been introduced to enhance the accuracy of fault detection and identification in MGs. An example of this is the integration of deep neural network techniques with Discrete Wavelet Transform, as presented in [15], to overcome fault detection challenges. Another approach in [16] incorporates a decision tree (DT)-based method with fuzzy logic, serving as a knowledge-based scheme, to facilitate reliable decision-making. However, it is worth noting that these approaches exhibit high-performance capabilities but necessitate extensive and comprehensive training data to effectively address the complexity and unprecedented conditions encountered in MG operations.

Another interesting protection strategy belongs to the multi-agent systems using various agents. For instance, a multi-agent-based protection system is proposed in [17] using the current angle of both sides of a line. Determination of the required criteria and the need for extensive data from different points make the multi-agent-based methods more expensive and complex.

In addition, with the advent of advanced measurement, computation, and communication technologies, the integration of these devices into MG protection systems has become widespread. While local protection functions remain essential due to their swift performance and lack of communication failure risks, the implementation of a wide-area structure becomes imperative in order to gather data from multiple points and make supervisory and back-up decisions, thereby enhancing the reliability and security of protection systems in MGs.

In this regard, based on the data acquired from phasor measurement units (PMUs) in MGs, different protection applications are managed [18]. A coordinated protection and control system is achieved in [19] using the frequency data from PMUs. Moreover, a supervisory level is produced to enhance the security and selectivity of the local protection systems by authors in [20]. In another approach, the PMU data is used for the MG islanding detection and provides a corrective strategy for the misoperation prevention of the protection systems.

It is noteworthy that wide-area protection applications are accessible concerning the speed, simplicity, security of the communication systems, and accuracy of the approaches.

Accordingly, this paper proposes a simple, accurate, and fast wide-area backup protection for fault detection and identification using the intelligent electronic devices (IEDs), installed in limited locations. More precisely, the proposed method uses data from IEDs without the need for line parameter information. According to the proposed method, each non-IED bus is surrounded by IED buses through the distribution lines. After the fault detection step, the faulted line is identified without the need for information related to the branches connected to the non-IED bus, such as lines and load information.

The proposed method does not need for criterion definition for fault detection. Indeed, it provides a model-free approach and is robust against events such as load-change and line outage.

Accordingly, the contributions and features of the proposed method are listed as follows.

- Presentation of an approach for fault detection without the need for line parameters and load information;
- Determination of the faulted line without the need for line and load information;
- Determination of the line parameters during the load-increasing event;
- Prevention of the misoperation protection during non-fault conditions.

The rest of the paper is organized as follows. Section II explains the details of the proposed method. Simulation results verify the proposed method's performance in section III. Finally, the last section belong to the conclusion and discussion about the performance of the proposed method.

2 METHODOLOGY

A novel numerical measurement-based scheme is proposed in this study, utilizing voltage and current data obtained from IEDs to establish an efficient and dependable approach for fault detection and identification. The following sections elaborate on the detailed process of the proposed method.

2.1 IED Placement Strategy

The installation of IEDs follows a strategic approach wherein IED buses surround each non-IED bus through the distributed lines. This placement strategy is graphically represented in Figure 1. According to this strategy, each IED bus is connected to at least one non-IED bus.

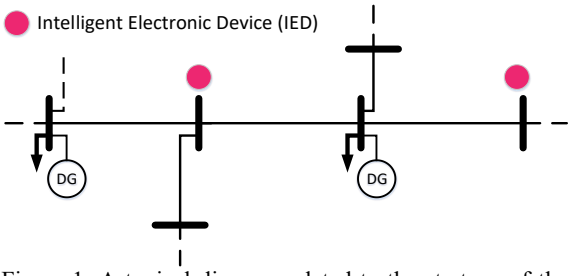


Figure 1: A typical diagram related to the strategy of the IEDs placement.

2.2 Fault Detection Approach

When a fault occurs at location f, the voltage at the fault location drops to zero, as depicted in Figure 2a. Consequently, the current flowing through the p-bus experiences a sudden increase due to the sharp voltage drop from the normal condition. To provide further clarity, the current drawn from the p-bus is calculated for both normal and fault conditions based on the Kirchhoff's Voltage Law (KVL), as outlined (1):

$$I_p = \frac{V_p - V_q}{Z_{pq}}, \qquad I_{p-f} = \frac{V_p - V_f}{Z_{pf}}. \tag{1}$$

where V_p, V_q, V_f, Z_{pq}, and Z_{pf} stand for the voltage of p, q, and fault location, the line impedance between p and q buses, and the line impedance between p and f location, respectively.

This observed increase in current during fault occurrence serves as an initial indicator for fault detection. It is important to note that this indication may also arise in similar events such as load-increasing and disconnection of DGs from non-IED buses. However, this rule can effectively differentiate faults from other events like line outages. During a line outage, the current in the associated line becomes zero. To distinguish fault occurrences from similar events like load-increasing and DG disconnection, the KVL is applied to the electrical circuit, as demonstrated in Figure 2b:

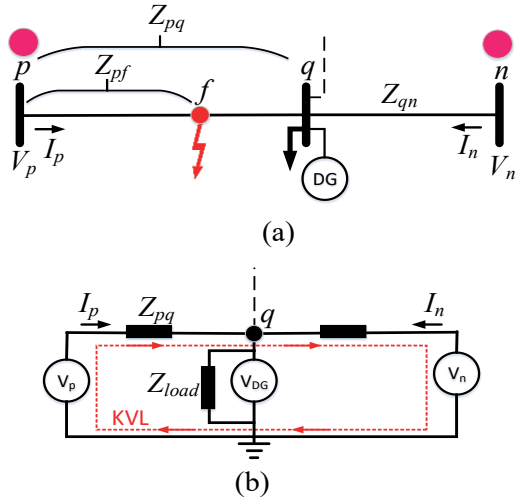


Figure 2: Schematic of the considered area: a) during the fault occurrence, b) equivalent circuit in normal condition.

$$V_p = Z_{pq} I_p + V_q - Z_{qn} I_n + V_n, \tag{2}$$

where Z_{qn}, I_n, and V_n are known as the line impedance between n and q buses, current flowing the line between n and q buses, and n-bus voltage, respectively. By dividing (2) to I_p, one can write:

$$\frac{V_p}{I_p} = Z_{pq} + \frac{V_q}{I_p} - Z_{qn}\frac{I_n}{I_p} + \frac{V_n}{I_p}. \tag{3}$$

It can be observed that the voltage and current values of the p and q buses are accessible through the installed IEDs, and (3) can be rearranged as follows.

$$Z_{pq} + \frac{V_q}{I_p} - Z_{qn}\frac{I_n}{I_p} + \frac{V_n}{I_p} - \frac{V_p}{I_p} = 0. \tag{4}$$

According to (4), there is an equation with two unknown variables, Z_{pq} and Z_{qn}.

In this step, when there is no fault or line outage, the values of Z_{pq} and Z_{qn} remain constant. Consequently, a numerical approach is employed to solve (4) and determine the unknown variables. To achieve this, two sequential samples of both voltage and current measurements are utilized, as follows (5):

$$\begin{aligned} Z_{pq} + \frac{V_q(h-1)}{I_p(h-1)} - Z_{qn}\frac{I_n(h-1)}{I_p(h-1)} + \frac{V_n(h-1)}{I_p(h-1)} - \frac{V_p(h-1)}{I_p(h-1)} = 0, \\ Z_{pq} + \frac{V_q(h)}{I_p(h)} - Z_{qn}\frac{I_n(h)}{I_p(h)} + \frac{V_n(h)}{I_p(h)} - \frac{V_p(h)}{I_p(h)} = 0, \end{aligned} \tag{5}$$

where h is the sample number of measurements.

Therefore, at each instance, there are two equations with two unknown variables. Based on the equations derived from the measurement samples, the values of Z_{pq} and Z_{qn} are calculated using the most

recent and previous samples. If the obtained values for these variables remain constant, it indicates the absence of faults on the considered lines for the given event. Conversely, if the values change, it signifies that one of the considered lines within the studied area has experienced a fault.

2.3 Fault Localization

Once a fault is detected, the next crucial step involves determining the specific location of the faulted line. Let us assume that the fault occurs at location f in Figure 3a, characterized by a resistance value of R. In Figure 3a, the variable x represents the ratio of the fault distance to the line length. To facilitate this process, the circuit's T-structure model is transformed into a Π model, as illustrated in Figure 3b.

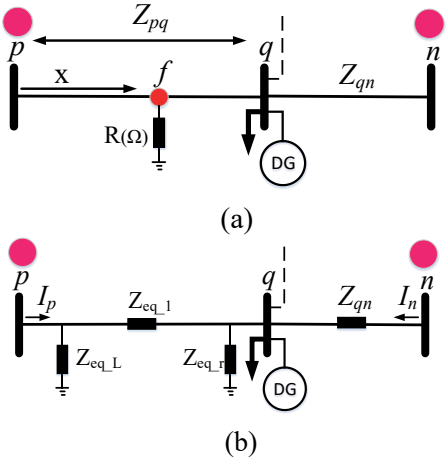


Figure 3: Diagram of the faulted area, a) location and resistance of the fault, b) transformed form of the faulted line.

By following process for the transformation of the delta connection to star connection based on the Kirchhoff's current Law (KCL) and KVL, the relationship is expressed as (6):

$$\begin{aligned} Z_{eq1} &= \frac{Z_1Z_2+Z_2R+Z_1R}{R}, \\ Z_{eq\_r} &= \frac{Z_1Z_2+Z_2R+Z_1R}{Z_1}, \\ Z_{eq\_l} &= \frac{Z_1Z_2+Z_2R+Z_1R}{Z_2}. \end{aligned} \tag{6}$$

Where:

$$\begin{aligned} Z_1 &= xZ_{pq}, \\ Z_2 &= (1-x)Z_{pq}. \end{aligned} \tag{7}$$

Applying the KVL and KCL to the circuit depicted in Figure 3b, we obtain (8) as follows:

$$V_n = Z_{nq}I_n - Z_{eq1}(I_p - \frac{V_p}{Z_{eq\_l}}) + V_p. \tag{8}$$

By simplification and dividing (8) to I_n, one can find:

$$\frac{V_n}{I_n} - Z_{nq} + Z_{pq}\frac{I_p}{I_n} - \frac{V_p}{I_n} = \frac{(1-x)Z_{pq}}{R}\frac{V_p}{I_n} - \frac{x(1-x){Z_{pq}}^2}{R}\frac{I_p}{I_n}. \tag{9}$$

In (9), we have two unknown variables, x and R. To solve this problem, a repetitive process is employed. For instance, assuming a numeric value for x, we calculate the corresponding value for R based on eq (9). If the obtained value for R is positive for one iteration and negative for the subsequent iteration of x, this pair of x and R is considered as the fault location and resistance on the relevant line. Otherwise, if the R is not a rational value, it indicates that the adjacent line has been subjected to the fault.

Figure 4 shows the overall flowchart of the proposed method for fault detection and identification.

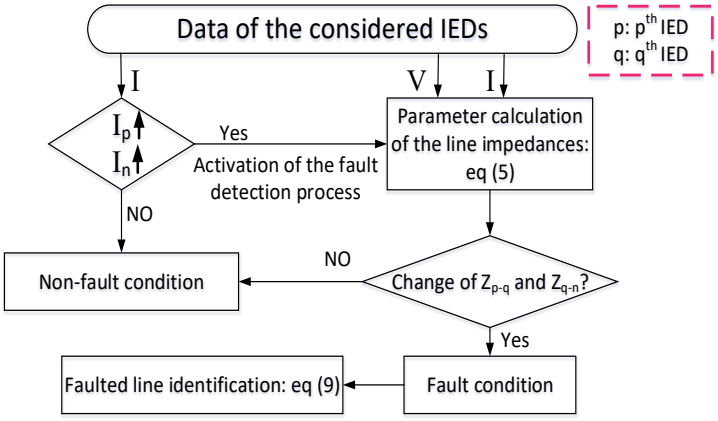


Figure 4: Flowchart of the proposed method for fault detection and faulted line determination.

3 SIMULATION AND RESULTS

In this section, simulation tests were performed to validate the proposed method within the MATLAB/SIMULINK environment. To this end, an 11-bus MG [21] was chosen as an autonomous test system to introduce the necessary disturbances. Figure 5 displays a schematic representation of the 11-bus test case.

For the initial evaluation, a three-phase fault with a resistance of R=0.001 Ω was applied to the middle of line 8-7, from t=0.1 s to t=0.2 s. In this scenario, the currents flowing through both lines, 8-7 and 7-3, increase at the moment of fault occurrence. The current trends are depicted in Figure 6a, indicating the activation of the fault detection process. The proposed approach of impedance calculation was employed in this regard. Figure 6b illustrates the results of the obtained $Z_{8\text{-}7}$ and $Z_{7\text{-}3}$ based on the proposed method. Notably, the actual impedance values are $Z_{8-7} = (0.0261 + 0.0025i)$ and $Z_{7-3} = (0.066 + 0.0128i)$. Notably, i stands for imaginary part of the complex numbers which is equal to $\sqrt{-1}$.

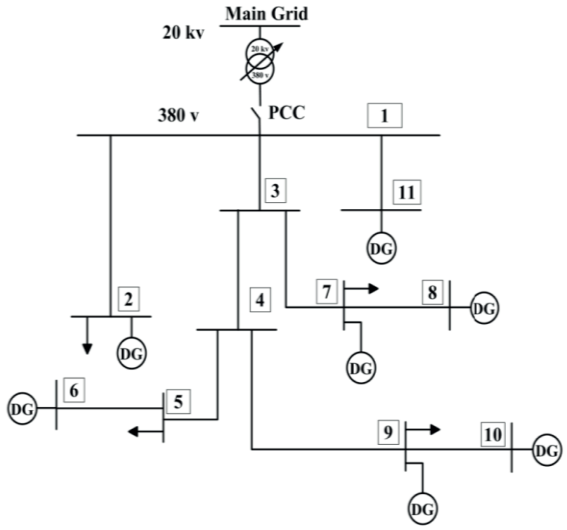


Figure 5: Single diagram of the 11-bus case study.

The obtained results demonstrate that the calculated impedances undergo changes during the fault occurrence, with magnitudes comparable to the actual impedance values. Following fault clearance, the calculated impedances return to their pre-fault conditions.

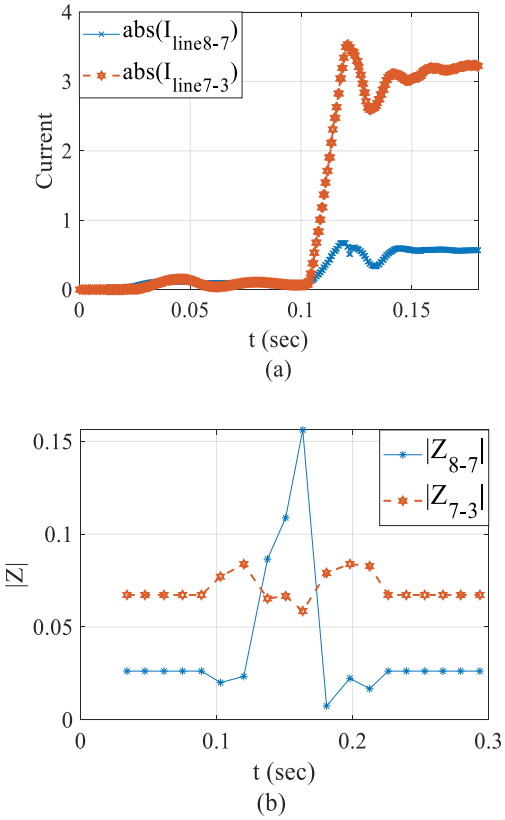


Figure 6: Simulation results for fault on line 8-7 with R=0.001 Ω: a) current values for both considered lines, b) calculated impedance.

To determine the faulted line, an iterative process involving line 8-7 and line 7-3 is initiated. Based on the obtained values of R, as indicated in Table 1, assuming the fault occurred on line 8-7, the calculated R values for different iterations of x are found to have positive real values. Specifically, the real part of the calculated R confirms that the fault was applied on line 8-7 with low resistance, and the fault location is likely to be within the range of x=0.1 to x=0.7. Conversely, assuming the fault occurred on line 7-3, the obtained R values, presented in Table 2, are negative and irrational.

Table 1: Calculated values of R for different x assuming the fault occurrence on line 8-7.

X	R
0.1	0.0043 - 0.0005i
0.3	0.0023 - 0.0003i
0.5	0.0010 - 0.0001i
0.7	0.0002 - 0.0001i
0.9	-0.0001 - 0.0000i

Assuming the fault occurrence on line 7-3, the calculated of R values for different fault distances (different x) will be negative, as shown in Table 2, indicating the false assumption of the fault location.

Table 2: Calculated values of R for different x assuming the fault occurrence on line 7-3.

x	R
0.1	-0.6881 - 0.7459i
0.3	-0.5043 - 0.5702i
0.5	-0.3381 - 0.4002i
0.7	-0.1896 - 0.2359i
0.9	-0.0588 - 0.0772i

To demonstrate the effect of load increase on the proposed method, the load connected to bus 7 was doubled compared to its normal condition. Despite exhibiting a similar trend to the currents with the fault conditions, the calculated impedance, as shown in Figure 7, remains unchanged. This observation indicates a non-fault condition.

In another scenario, the DG connected to bus 7 was disconnected from t=0.1 s to t=0.2 s. The results indicate that the calculated impedances for $Z_{8\text{-}7}$ and $Z_{7\text{-}3}$ remain constant, as depicted in Figure 8. Notably, during this condition, conventional protection systems may make incorrect decisions due to the increased current. However, the proposed method can easily discriminate between fault and non-fault conditions.

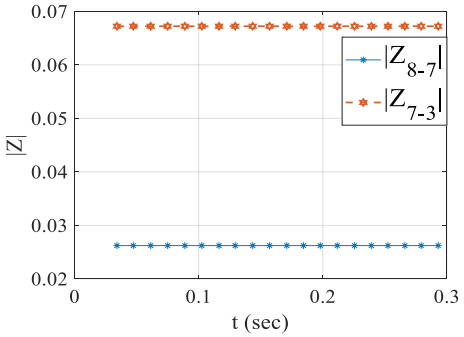


Figure 7: Absolute value of the calculated impedance for the load-increasing from t=0.1 s to t=0.2 s.

Another challenging event for conventional protection systems is line outage. To evaluate the performance of the proposed method, line 8-7 was temporarily removed from the system from t=0.1 s for a duration of 0.1 s. As shown in Figure 9, the line current becomes zero, while the magnitude of the calculated impedance parameters changes significantly, indicating a line outage. The decrease in current and high values of the impedance parameters imply a non-fault condition.

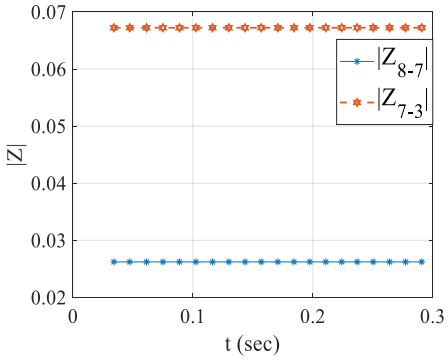
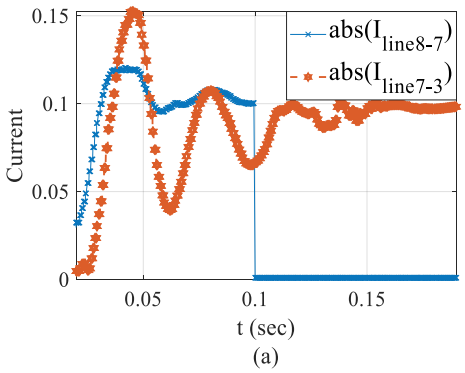


Figure 8: Absolute value of the calculated impedance for DG outage from bus 7 from t=0.1 s to t=0.2 s.



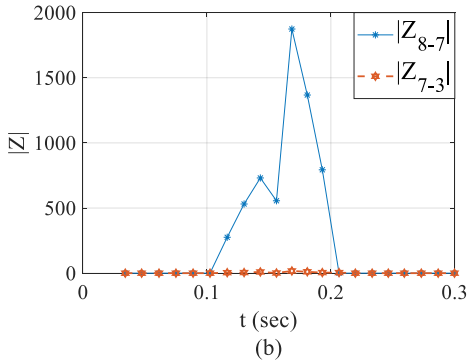


Figure 9: Results in response to line 8-7 outage at t=0.1 s to t=0.2 s, a) current trajectory of the considered lines, b) absolute values of the calculated impedances.

For further investigations, a three-phase fault with R=50 Ω was applied to the middle of line 8-7 at t=0.1 s, lasting for 0.1 s. The calculated impedances, as shown in Figure 10, confirm the occurrence of a fault at t=0.1 s, based on the changing values of the calculated parameters.

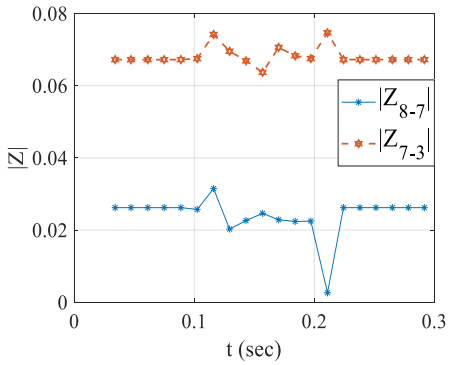


Figure 10: Absolute values of the calculated impedances for fault on the line 8-7 with R=50 Ω.

In another scenario, line 9-4 was subjected to a three-phase fault with R=10 Ω. The fault was located at a distance 0.2 times the line length from bus 9 and lasted for 0.1 s from t=0.1 s. The calculated impedances, $Z_{9\text{-}4}$ and $Z_{5\text{-}4}$, representing the considered lines impedances, are presented in Figure 11. The obtained results demonstrate that the calculated parameters change simultaneously with the fault occurrence, starting from t=0.1 s.

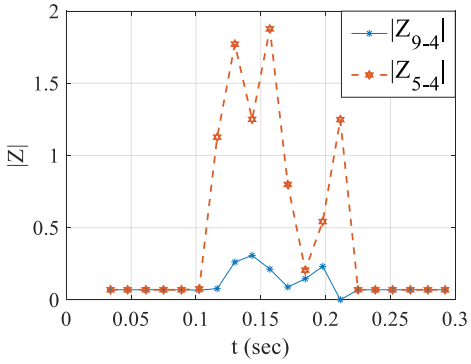


Figure 11: Absolute values of the calculated impedances for the fault on the line 9-4 from t=0.1 s to t=0.2 s with R=10 Ω.

Table 3 displays the obtained parameters related to the fault resistance for x=0.2, assuming the fault occurred on line 9-4. The positive value confirms the assumption of fault location. Conversely, assuming the fault occurred on line 5-4, the obtained R values are not acceptable. According to the obtained values for R assuming the fault occurrence on lines 9-4 and 5-4, respectively, in Table 3, the actual fault location is determined based on the rational and positive R value when line 9-4 has been subjected to the fault, especially for x=0.2.

Table 3. Values of R for x=0.2 assuming fault occurrence on the lines 9-4 and 5-4.

Assuming the fault on the line 9-4	
x	R
0.2	9.8910 + 0.3160i
Assuming fault on the line 5-4	
x	R
0.2	-9.9486 - 0.3086i

The effect of the applied fault on line 9-4 was investigated by analyzing the calculated impedances of $Z_{5\text{-}4}$ and $Z_{3\text{-}4}$, representing the main lines impedances. The results, as depicted in Figure 12, indicate that the values of $Z_{5\text{-}4}$ and $Z_{3\text{-}4}$ remain constant throughout the fault occurrence on line 9-4, which is connected to bus 4. This observation verifies that faults occurring on lines outside the considered area have no effect on the calculated impedances, and they remain unchanged.

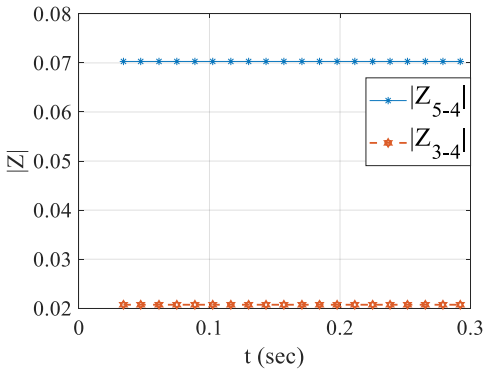


Figure 12: Absolute of calculated impedances related to $Z_{5\text{-}4}$ and $Z_{3\text{-}4}$ during the fault on the line 9-4 at t=0.1 s to t=0.2 s.

In this work, the reporting rate of IEDs was set to 60 samples per second. Consequently, the decision-making time for fault detection, assuming two samples for each decision, was determined to be 34 ms. This fast decision-making time demonstrates the efficient performance of the proposed method.

4 CONCLUSIONS

This paper presents an innovative data-driven wide-area backup protection approach for fault detection. The proposed method offers a reliable solution for detecting fault conditions along the primary route of the lines between the two IED-buses, without relying on detailed information about the lines, load, DGs, or the equipment connected to non-IED-buses. Through a repetitive approach, the faulted line is accurately identified following fault detection.

The results obtained from this study reveal that the correct performance of the proposed method remains unaffected by non-fault disturbances. Furthermore, the efficacy of the approach remains intact even when faults occur outside the designated area of interest. These findings serve as concrete evidence, validating the speed of the proposed method, which requires a maximum of two sequential calculated samples to arrive at a suitable decision.

It is important to note that the potential impact of measurement errors on the performance of the proposed method has not been taken into consideration. Additionally, despite the limited number of measurements employed in this paper, future research endeavors will explore an evolved method that relies on a reduced number of measurements, aiming to enhance fault detection capabilities.

REFERENCES

[1] M. Farrokhabadi et al., "Microgrid Stability Definitions, Analysis, and Examples," IEEE Trans. Power Syst., vol. 35, no. 1, pp. 13-29, 2020, doi: 10.1109/TPWRS.2019.2925703.

[2] W.M. Hamanah, M.I. Hossain, M. Shafiullah, and M.A. Abido, "AC Microgrid Protection Schemes: A Comprehensive Review," IEEE Access, vol. 11, pp. 76842-76868, 2023, doi: 10.1109/ACCESS.2023.3298306.

[3] G.S. Dua, B. Tyagi, and V. Kumar, "Fault Detection Technique for Distribution Networks and Microgrids Using Synchrophasor Data," IEEE Trans. Ind. Appl., pp. 1-14, 2023, doi: 10.1109/TIA.2023.3305362.

[4] M.W. Altaf, M.T. Arif, S.N. Islam, and M.E. Haque, "Microgrid Protection Challenges and Mitigation Approaches–A Comprehensive Review," IEEE Access, vol. 10, pp. 38895-38922, 2022, doi: 10.1109/ACCESS.2022.3165011.

[5] P.H.A. Barra, D.V. Coury, and R.A. S. Fernandes, "A survey on adaptive protection of microgrids and distribution systems with distributed generators," Renew. Sustain. Energy Rev., vol. 118, September 2019, p. 109524, 2020, doi: 10.1016/j.rser.2019.109524.

[6] A. Dagar, P. Gupta, and V. Niranjan, "Microgrid protection: A comprehensive review," Renew. Sustain. Energy Rev., vol. 149, no. June, p. 111401, 2021, doi: 10.1016/j.rser.2021.111401.

[7] S. Mirsaeidi, D. Mat Said, M.W. Mustafa, and M. Hafiz Habibuddin, "A protection strategy for micro-grids based on positive-sequence component," IET Renew. Power Gener., vol. 9, no. 6, pp. 600–609, 2015, doi: https://doi.org/10.1049/iet-rpg.2014.0255.

[8] K.A. Saleh and A. Mehrizi-Sani, "Harmonic Directional Overcurrent Relay For Islanded Microgrids With Inverter-Based DGs," IEEE Syst. J., vol. 15, no. 2, pp. 2720-2731, 2021, doi: 10.1109/JSYST.2020.2980274.

[9] A.D. Bebars, A.A. Eladl, G.M. Abdulsalam, and E.A. Badran, "Internal electrical fault detection techniques in DFIG-based wind turbines: a review," Prot. Control Mod. Power Syst., vol. 7, no. 1, p. 18, 2022, doi: 10.1186/s41601-022-00236-z.

[10] M.N. Alam, "Overcurrent protection of AC microgrids using mixed characteristic curves of relays," Comput. Electr. Eng., vol. 74, pp. 74-88, 2019, doi: https://doi.org/10.1016/j.compeleceng.2019.01.003.

[11] P.T. Mandi tereza and R.C. Bansal, "Protection of microgrids using voltage-based power differential and sensitivity analysis," Int. J. Electr. Power Energy Syst., vol. 118, p. 105756, 2020, doi: https://doi.org/10.1016/j.ijepes.2019.105756.

[12] A.N. Sheta, G.M. Abdulsalam, B.E. Sedhom, and A.A. Eladl, Comparative framework for AC-microgrid protection schemes: challenges, solutions, real applications, and future trends, vol. 8, no. 1. Springer Nature Singapore, 2023. doi: 10.1186/s41601-023-00296-9.

[13] N. Hussain, M. Nasir, J.C. Vasquez, and J. M. Guerrero, "Recent developments and challenges on AC microgrids fault detection and protection systems-a review," Energies, vol. 13, no. 9, 2020, doi: 10.3390/en13092149.

[14] J. C. Vasquez, "Adaptive protection combined with machine learning for microgrids," IET Gener. Transm. Distrib., vol. 13, no. 6, pp. 770-779(9), March 2019, [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/iet-gtd.2018.6230

[15] J.J.Q. Yu, Y. Hou, A.Y.S. Lam, and V.O.K. Li, "Intelligent Fault Detection Scheme for Microgrids With Wavelet-Based Deep Neural Networks," IEEE Trans. Smart Grid, vol. 10, no. 2, pp. 1694-1703, 2019, doi: 10.1109/TSG.2017.2776310.

[16] S. Kar and S. R. Samantaray, "A Fuzzy Rule Base Approach for Intelligent Protection of Microgrids," Electr. Power Components Syst., vol. 43, no. 18, pp. 2082-2093, 2015, doi: 10.1080/15325008.2015.1070384.

[17] F.A.S. Dizioli, P.H.A. Barra, T.S. Menezes, V.A. Lacerda, D.V Coury, and R.A.S. Fernandes, "Multi-agent system-based microgrid protection using angular variation: An embedded approach," Electr. Power Syst. Res., vol. 220, p. 109324, 2023, [Online]. Available: https://doi.org/10.1016/j.epsr.2023.109324.

[18] R. Jain, Y. Nag Velaga, K. Prabakar, M. Baggu, and K. Schneider, "Modern trends in power system protection for distribution grid with high DER penetration," e-Prime - Adv. Electr. Eng. Electron.

Energy, vol. 2022, no. July, pp. 2772-6711, 2022, doi: 10.1016/j.prime.2022.100080.

[19] Y. Seyedi and H. Karimi, "Coordinated Protection and Control Based on Synchrophasor Data Processing in Smart Distribution Networks," IEEE Trans. Power Syst., vol. 33, no. 1, pp. 634-645, 2018, doi: 10.1109/TPWRS.2017.2708662.

[20] P. Kundu and A.K. Pradhan, "Supervisory protection of islanded network using synchrophasor data," IEEE Trans. Smart Grid, vol. 10, no. 2, pp. 1772-1780, 2019, doi: 10.1109/TSG.2017.2777873.

[21] S. Papathanassiou, "Study-Case LV Network," [Online]. Available: http://microgrids.power.ece.ntua.gr/documents/Study-Case%20LV-Network.pdf.

Robust Tuning of Grid-Forming Converters Using Kharitonov Theorem

Sharara Rehimi[1], Hassan Bevrani[2], Chiyori Urabe[1], Takeyoshi Kato[1] and Toshiji Kato[3]
[1]*Electrical Engineering Department, Institute of Materials and Systems for Sustainability, Nagoya University, Tsurumaicho Str. 65, Showa Ward, Nagoya, Japan*
[2]*Electrical Engineering Department, Smart/Micro Grids Research Center, University of Kurdistan, Pasdaran Boulevard, PO Box 416, Sanandaj, Iran*
[3]*Electrical Engineering Department, Doshisha University, Gembucho Str. 601, Kamigyo-ku, Kyoto, Japan*
rehimi.sharara.z8@s.mail.nagoya-u.ac.jp, bevrani@uok.ac.ir, urabe.chiyori.c0@f.mail.nagoya-u.ac.jp, kato.takeyoshi.b5@f.mail.nagoya-u.ac.jp, tkato@mail.doshisha.ac.jp

Keywords: Kharitonov Theorem, Grid-Forming Converters, Robust Tuning, Modern Power Grids.

Abstract: This paper presents a study on the robust tuning of damping coefficient, inertia, and conventional controller gains for a grid-forming converter connected to the main grid. The proposed method is evaluated using the Typhoon Hill system, which allows for comprehensive simulation and analysis. The Kharitonov theorem is utilized to achieve robustness in the face of deviations in line parameters. Unlike previous works, our approach systematically and graphically searches for a non-conservative Kharitonov region in the solution area of the controller coefficients. This region characterizes all stabilizing gain controllers that effectively stabilize an uncertain control structure. By selecting coefficients from the obtained non-conservative Kharitonov region, the synthesized controller effectively stabilizes the grid-forming converter. The results of this study highlight the efficacy of the Kharitonov theorem in achieving robust tuning of essential parameters for grid-forming converters, enhancing stability and performance in the presence of line parameter variations.

1 INTRODUCTION

1.1 Modern Power Grids

In recent years, there has been a notable increase in the integration of renewable energy sources (RESs) into the modern power grids, driven by the demand for sustainable and environmentally friendly energy solutions. Power electronic interface converters (grid connected converters) are essential for connecting RESs to the grid, enhancing controllability and enabling RESs to contribute effectively to grid objectives. Researchers have focused on improving the efficiency of these converters, leading to significant advancements in power converter technology [1].

Grid connected converters can be categorized as grid forming converters (GFM) and grid following converters (GFL). The GFM converters, acting as voltage sources, offer greater flexibility compared to GFL converters, which function as current sources. Various control methods have been developed to fully utilize the capabilities of GFM converters. In addition to connecting RESs to the grid, GFM converters can provide ancillary services such as frequency adjustment, voltage regulation, improved inertial response, damping enhancement, and power factor correction [2].

The significance of GFM converters lies in their ability to meet the diverse needs of modern power grids. Power companies, particularly those heavily reliant on RESs, are revising their grid codes to leverage the capabilities of GFM converters and optimize their support to the main grid. This strategic response allows for enhanced integration of RESs and maximizes grid interaction, positioning GFM converters as integral interfaces in the modern power systems [2].

Robust tuning of control coefficient in the control system of a GFM is a crucial aspect of ensuring stable and reliable operation in modern power grids. The GFMs play a vital role in maintaining grid stability and providing ancillary services. However, uncertainties in line parameters and operating conditions can affect the performance of the converter control system. To address this challenge, researchers

have focused on robust tuning of control coefficients in the control system of the GFMs [3].

1.2 Robust Control Approach

Robust tuning involves selecting appropriate values for the most important control gains in the GFM control system to ensure system stability and resilience in the face of parameter variations. The Kharitonov theorem has emerged as a powerful tool in this context [4]. It allows for a systematic and graphical search for a non-conservative Kharitonov region in the controller coefficient parameter space. This region represents all stabilizing Proportional-Integral-Derivative (PID) controllers that can effectively stabilize an uncertain control system. By selecting coefficients from this region, the synthesized controller can effectively stabilize the GFM.

As mentioned, the Kharitonov theorem is a key tool used in several studies to design robust controllers for uncertain control systems. A study in [5] provides an elementary proof of the Kharitonov theorem based on Bezoutian matrices. This proof establishes a unified derivation of Kharitonov-like theorems for both continuous-time and discrete-time settings. The study introduces the concept of (block) Anderson-Jury Bezoutians and suggests a potential technique to address a robust stability problem in the multi input-multi output (MIMO) case. In [6], the theorem is applied to synthesize stabilizing controllers for interval plants. The study focuses on designing a controller that can stabilize multiple Kharitonov-defined vortex polynomials simultaneously. Through the search of a non-conservative Kharitonov region, the controller is systematically developed within the bounds of all stabilizing PID controllers for uncertain plants. The study also incorporates a virtual gain phase margin tester compensator to ensure robust safety margins.

Authors in [7] propose a tuning strategy for robust static output feedback (SOF) controllers using Kharitonov's theorem. The goal is to minimize the quadratic cost of the controllers while considering multiple parametric uncertainties. The study utilizes Kharitonov's theorem to define a family of bounded, robustly stable SOF controllers. An evolutionary algorithm is then employed to select the controller that minimizes the quadratic cost within this family. This approach addresses the computational complexity of the control problem's non-convex nature. In [8], a robust decentralized proportional-integral (PI) control design for load frequency control (LFC) in a multi-area power system is presented. The study optimizes the system's robustness margin and transient performance simultaneously by determining the PI controller parameters using Kharitonov's theorem. The theorem helps establish the maximal uncertainty bounds for stable power system performance, enabling the optimization of control objectives through techniques like genetic algorithms.

In the field of LFC, another study focuses on efficient control algorithms [9]. The paper explores the robustness of control algorithms based on fractional order control and interval modeling of the plant. By utilizing Kharitonov's theorem, a fractional order PID controller is designed for the interval model of a single area LFC. This approach allows engineers to define parameter limits that ensure satisfactory controller performance, leading to improved system stability and safety. Furthermore, these studies demonstrate the versatility and usefulness of the Kharitonov theorem in various control design applications, ranging from interval plants to power systems. The theorem enables the synthesis of stabilizing controllers, optimization of control objectives, and robustness analysis of control algorithms.

Meanwhile, authors in [10] present a method for designing robust power system stabilizers (PSSs). The stabilizing PSSs are characterized using d-decomposition, which divides the controller-parameter space into root-invariant regions. The designed control is applied to a single machine-infinite bus system (SMIB). The approach considers d-stability, which incorporates a pre-specified damping cone in the complex s-plane to improve time-domain specifications. The pole clustering in the damping cone is achieved by ensuring the Hurwitz stability of a complex polynomial. The parametric uncertainties caused by variations in load patterns are captured using an interval polynomial. The computation of the set of robust d-stabilizing PSSs requires checking the Hurwitz stability of a complex interval polynomial, which is addressed using a complex version of Kharitonov's theorem.

Besides, a study in [11], focuses on LFC of a microgrid in the presence of cyber-attacks. The paper proposes the application of the Kharitonov theorem-based proportional-integral (KT-PI) control technique. The microgrid consists of RESs, controllable energy sources, and energy storage systems. A denial-of-service (DoS) attack is considered as the cyber-attack scenario. The robustness and effectiveness of the KT-PI control technique are evaluated by considering parametric uncertainties and nonlinearities such as generation

rate constraint (GRC) and governor dead-band (GDB). Performance indices such as integral square error (ISE), integral absolute error (IAE), and integral of time absolute error (ITAE) are used for quantitative comparative analysis against the ZN-PI controllers. Then, the stability of the microgrid under the DoS attack is analyzed using the eigenvalue approach.

Another work addresses the robust stability analysis of an islanded microgrid with droop-controlled inverter-based distributed generators (DGs) [12]. The low-frequency (LF) dominant modes of the microgrid can become unstable or highly oscillatory due to large load changes, microgrid structure reconfiguration, and higher-power demands. To address this, a robust two-degree-of-freedom (2DOF) decentralised droop controller is proposed for each DG unit. A new design procedure is presented to robustly determine the transient droop gains, which effectively damp the LF oscillatory modes of the microgrid in the presence of disturbances, equilibrium point variations, and uncertain parameters. Inspired by Kharitonov's stability theorem, a robust D-stability analysis is performed to determine the specific ranges of the transient droop gains.

Furthermore, another study is focused on the stability analysis of the pitch system control of a horizontal axis wind turbine using the Kharitonov robust stability method. The objective is to assess the robust stability of the pitch controller under uncertainties arising from varying operating conditions. The study utilizes the National Renewable Energies Laboratory (NREL) 5 MW class IIA wind turbine model. The proposed method demonstrates satisfactory response to limited variations in the models characteristics [13].

Additionally, a reduction method for higher-order interval systems based on Kharitonov's theorem is presented in [14]. The method utilizes differentiation to achieve a reduced-order interval model. It is applicable to both single-input single-output (SISO) and MIMO interval systems. The proposed method is mathematically simple and preserves the dominant characteristics of the original system in its reduced-order interval models. Stable reduced-order models are obtained if the original system is stable. The effectiveness of the technique is demonstrated through benchmark problems, showing close approximation to the original system. The proposed method is compared with other well-known reduction methods in terms of time response specifications and performance indices, validating its effectiveness and efficiency. Additionally, the method is extended for discrete-time interval systems.

The present study focuses on the extensive and impactful utilization of Kharitonovs theorem for robust coefficient adjustment of a GFM control systems. Specifically, the study applies this method to adjust the coefficients of the key parameters in the control system of GFM.

This paper is organized as follows: Section 2 presents the dynamic modeling of the case study. Section 3 describes the mathematical model used for performance evaluation of the proposed robust control method. In Section 4, the control design procedure based on Kharitonovs theorem is presented. Section 5 presents the simulation results, demonstrating the effectiveness of the proposed control strategy. Finally, Section 6 concludes the paper by summarizing the key findings and contributions of this study.

2 DYNAMIC MODELING

The GFM under investigation is connected with an LCL filter. Subsequently, the system is linked to the line impedance and then connected to the main power grid as shown in Figure 1. Here, the L_{f1}, R_{f1}, C_f, R_c, L_{f2}, R_{f2}, L_l, R_l and V_{DC} represent the filter inductor (left side), filter inductor ohmic resistance (left side), filter capacitor, filter capacitor ohmic resistance, filter inductor (right side), filter inductor ohmic resistance (right side), line inductor, line ohmic resistance, and DC voltage link, respectively. Furthermore, the control system is implemented on the GFM converter using pulse width modulation (PWM). Additional information regarding the control system unit is illustrated in Figure 2.

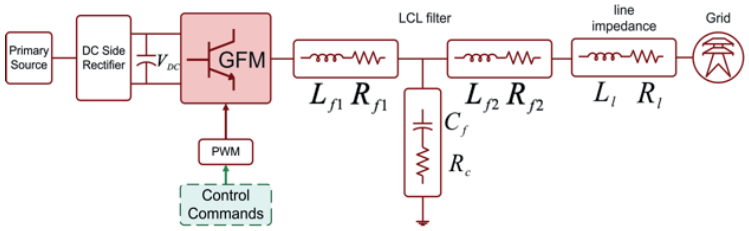


Figure 1: Grid connected converter structure: grid forming mode.

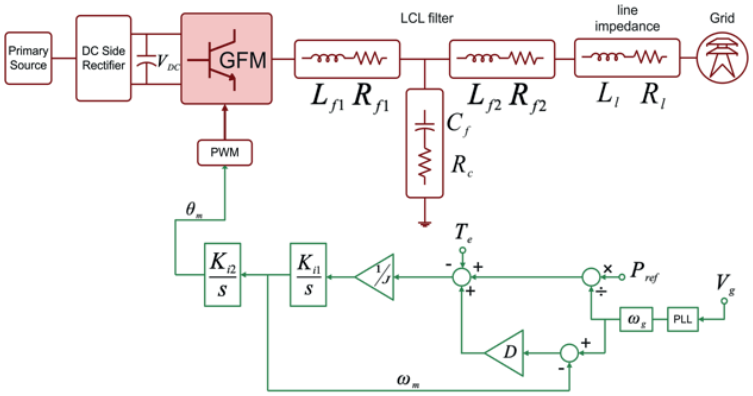


Figure 2: Active power control system of a grid forming converter.

.

The control system utilized in this paper includes active power control, as illustrated in Figure 2. Here, T_e, PLL, ω_g, P_{ref}, ω_m, J, K_{i1}, K_{i2}, D, V_g, and θ_m are electrical torque, phase-locked loop (PLL) dynamic (in this study considered 1), grid frequency, reference active power, measured frequency, inertia coefficient, first conventional control gain, second conventional control gain, damping coefficient, grid voltage and measured angle to be applied to the PWM, respectively.

3 MATHEMATIC MODEL

This section presents the mathematical model of the studied system, specifically focusing on the power section (LCL filter and line section) as shown in Figure 1. The main objective is to derive the transfer function (1) that maps the input voltage to the corresponding output voltage [1].

$$G_p = G_{Filter} G_{Line} . \quad (1)$$

Where G_p, G_{Filter}, and G_{Line} are the power section, LCL filter, and line transfer function, respectively. G_p can be calculated (2) as follows:

$$G_p = \frac{(A)s^6 + (B)s^5 + (C)s^4 + (H)s^3 + (F)s^2 - (E)s - (1)D}{(K)s^5 + (L)s^4 + (O)S^3 + (N)S^2 + (M)s} . \quad (2)$$

Once the transfer function of the power section in the studied system is obtained, the characteristic equation can be derived based on the obtained transfer function (3):

$$\Delta(s) = 1 + G_{Filter} G_{Line} K(s) = 0 , \quad (3)$$

where:

$$K_s = \frac{(Y)s}{(Z)s^2 + (R)s} , \quad (4)$$

and

$$\Delta(s) = a_6 s^6 + a_5 s^5 + a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s^1 + a_0 - D . \quad (5)$$

The detailed parameters are available in the appendix section. Additionally, system parameters are given in Table 1.

Table 1: System parameters.

Parameter	Value	Parameter	Value
T_{fP}	0.0159 s	L_f	0.15pu
T_{fQ}	0.0157 s	R_f	0.005pu
ω_0	377 rad/s	L_l	0.15 pu
L_v	0.32pu	C_f	0.064 pu
R_v	0.03 pu	R_c	0.002 pu
S_{base}	5 kVA	v_0	200 V

4 PROBLEM STATEMENT

Optimizing the conventional controller gains within the control unit of a GFM is a challengeable task while considering uncertain conditions encountered in practical sites. KT offers a solution for assessing the stability of systems that possess uncertain parameters. The theorem is employed by examining the extreme scenarios of coefficient ranges to determine if the system remains stable across all possible coefficient values. In this study, KT is applied to optimize the parameters of GFM control unit under the uncertainties condition. To achieve this, the important control gains are computed based on the calculation of the Kharitonov polynomials. The controller parameters are selected from a specific region that exhibits the best behavior in the root locus plan.

5 CONTROL DESIGN PROCEDURE

The mathematical model of the system under consideration, which incorporates parametric uncertainties, is represented as follows (6):

$$\Delta(s) = [a_6^+, a_6^-]s^6 + [a_5^+, a_5^-]s^5 + [a_4^+, a_4^-]s^4 + [a_3^+, a_3^-]s^3 + [a_2^+, a_2^-]s^2 + [a_1^+, a_1^-]s^1 + [a_0^+, a_0^-] - D \quad (6)$$

Where a_i^+ and a_i are the lower and upper bounds of a_i, respectively. The polynomials for KT can be written as:

$$K1 = a_0^- + a_1^- s + a_2^+ s^2 + a_3^+ s^3 + a_4^- s^4 + a_5^- s^5 + a_6^+ s^6,$$
$$K2 = a_0^+ + a_1^+ s + a_2^- s^2 + a_3^- s^3 + a_4^+ s^4 + a_5^+ s^5 + a_6^- s^6,$$
$$K3 = a_0^- + a_1^+ s + a_2^+ s^2 + a_3^- s^3 + a_4^- s^4 + a_5^+ s^5 + a_6^+ s^6,$$
$$K4 = a_0^+ + a_1^- s + a_2^- s^2 + a_3^+ s^3 + a_4^+ s^4 + a_5^- s^5 + a_6^- s^6.$$

Considering 20% variation in the line parameters $(R_l - L_l)$, a_i^+ and a_i can be obtained and the polynomials for KT can be calculated. This study focuses on the adjustment of the inertia coefficient, damping coefficient, and conventional controller coefficient through robust-based techniques. By considering positive values for the inertia and damping coefficients, the admissible range of values for the target coefficients is determined to ensure stability for all Kharitonov polynomials. The stability regions corresponding to each polynomial are shown in Figure 3. The intersection of these regions represents the feasible solution area according to KT. The common solution area is illustrated in Figure 4. The basic geometry associated with the zero-exclusion condition, for $0 < \omega < 50$ kHz, is fully demonstrated in Figure 5. Based on the zero-exclusion condition closed-loop system with designed controller is robustly stable if and only if, the rectangle plots do not include the origin of plane. This issue is clearly confirmed in Figure 5.

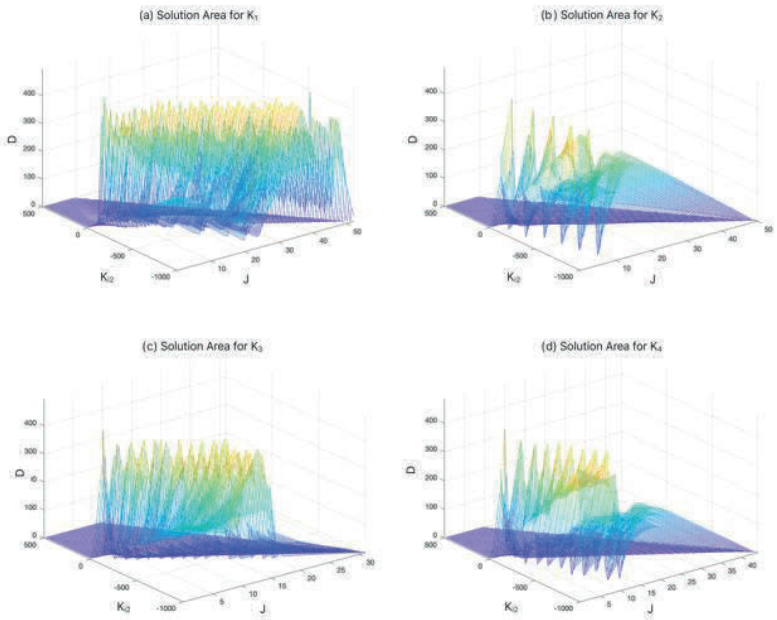


Figure 3: Stability regions corresponding to each polynomial: a) first KT polynomial, b) second KT polynomial, c) third KT polynomial, d) fourth KT polynomial.

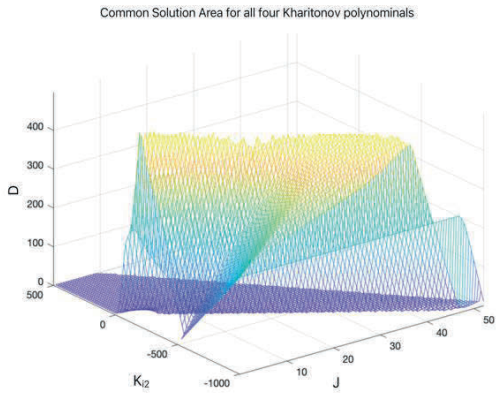


Figure 4: The common solution area for all four KT polynomials.

6 SIMULATION RESULTS

This section focuses on the evaluation of the controller gain values robustly tuned for active power control in a GFC control system. The Typhoon Hill software is employed to simulate and assess the effects of these proposed controller gains on the system. Figure 6 presents the configuration of the system under investigation within the Typhoon HIL software. Result illustrated in Figure 7, the system performance under normal conditions Figure 7a is depicted as being stable and acceptable. However, when a 20% variation is applied to the impedance values of the interconnecting line between the filter and the grid, the output power experiences fluctuations Figure 7b. By employing the proposed method to adjust the existing controller gains, the system behavior is restored to its normal state, thus achieving stability Figure 8c.

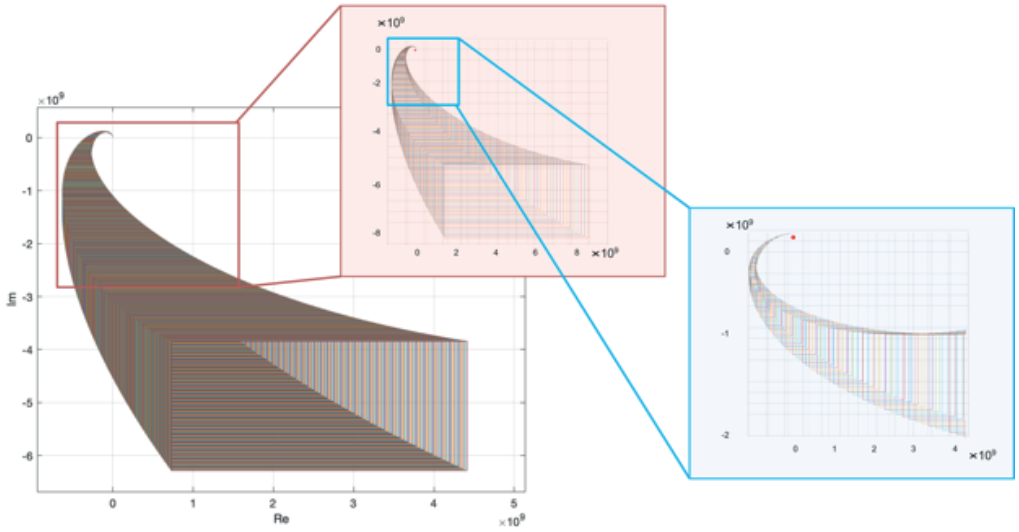


Figure 5: Motion of the Kharitonov's rectangle for 0 < ω< 50 kHz.

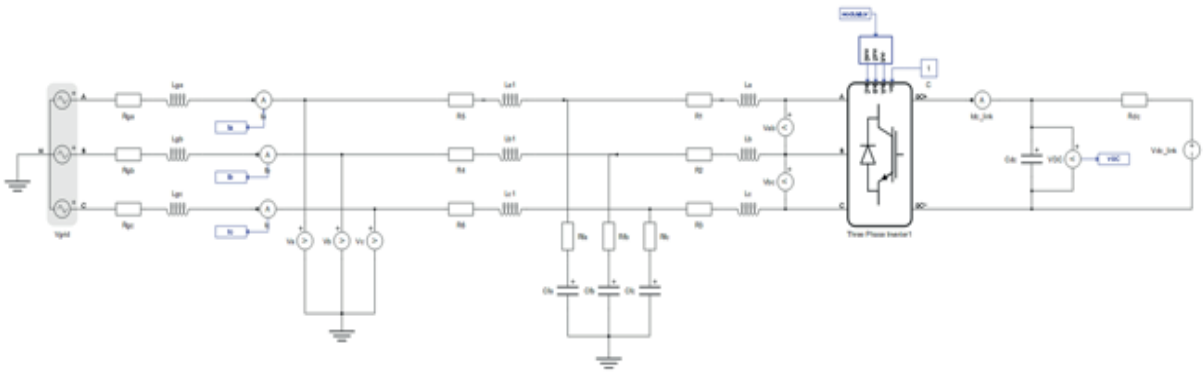


Figure 6: Grid forming converter: Typhoon HIL structure.

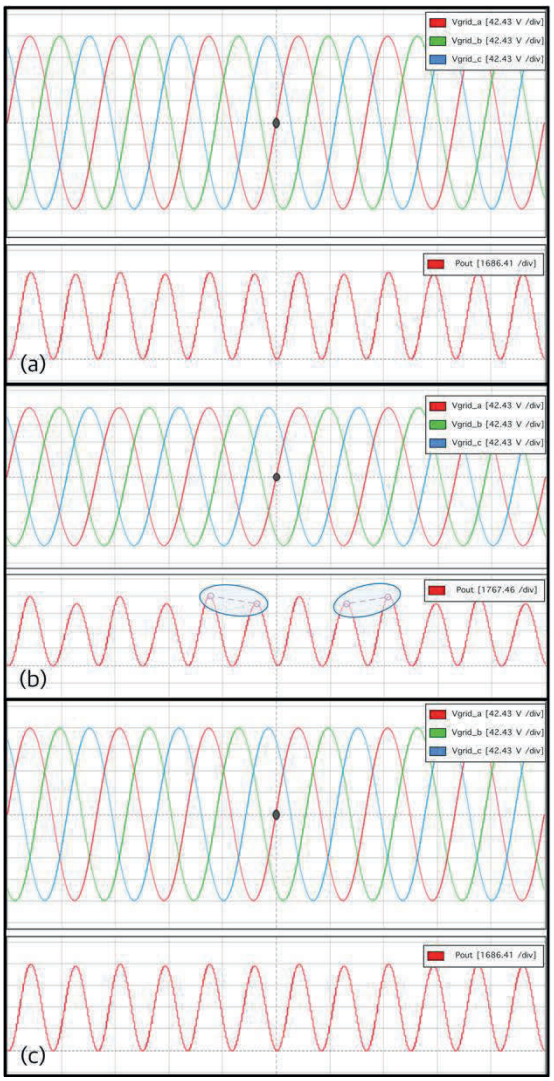


Figure 7: Voltage and active power: a) normal condition, b) 20% deviations in the line parameters with normal controller, c) 20% deviations in the line parameters with robust tuned controller.

7 CONCLUSIONS

This research employs Kharitonovs theory to determine the acceptable performance range of a grid-forming converter under conditions of uncertainty in the transmission lines parameters. By utilizing this methodology, the need for additional controllers to maintain system stability in the presence of uncertainty is eliminated. Instead, the existing conventional controllers can be re-tuned using this theory to effectively address situations involving uncertainty.

Indeed, this study demonstrates the effectiveness of the robust tuning of damping coefficient, inertia, and conventional controller gains in a grid-forming converters. By Using the Kharitonov theorem and utilizing the Typhoon Hill device for evaluation, the stability and enhanced performance in the presence of line parameter variations have successfully achieved. Through systematic and graphical exploration of the non-conservative Kharitonov region, authors have identified stabilizing existing controllers that effectively stabilize uncertain control structures. The findings of this study contribute to the field by providing a comprehensive approach to robust parameter tuning, ultimately strengthening the stability and reliability of grid-forming converters in practical applications.

ACKNOWLEDGMENTS

The authors acknowledge the financial support of KDDI Foundation, Japan. Also received support from Smart/Micro Grids Research Center (SMGRC) at the University of Kurdistan, Sanandaj, Iran is highly appropriated.

REFERENCES

[1] F. Blaabjerg, Y. Yang, K. A. Kim, and J. Rodriguez, "Power electronics technology for large-scale renewable energy generation," Proceedings of the IEEE, vol. 111, no. 4, pp. 335-355, Mar. 2023.

[2] H. Bevrani et al., "Grid Connected Converters, Modeling, Stability and Control," Elsevier, Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands, 2022.

[3] H. Bevrani, "Robust Power System Frequency Control," 2nd ed., J. H. Chow, A. M. Stankovic, and D. Hill, Eds. Springer, Cham Heidelberg New York Dordrecht London, 2014.

[4] B. R. Barmish, "New Tools for Robustness of Linear Systems," Macmillan Coll Div, 1993.

[5] A. Olshevsky and V. Olshevsky, "Kharitonov's Theorem and Bezoutians," Linear Algebra and its Applications, vol. 399, pp. 285-297, 2005.

[6] D. G. Padhan and S. Majhi, "Enhanced cascade control for a class of integrating processes with time delay," ISA Trans, vol. 52, no. 1, pp. 45-55, 2013.

[7] R. Toscano and P. Lyonnet, "Robust static output feedback controller synthesis using Kharitonov's theorem and evolutionary algorithms," Information Sciences, vol. 180, no. 10, pp. 2023-2028, 2010.

[8] M. R. Toulabi, M. Shiroei, and A. M. Ranjbar, "Robust analysis and design of power system load frequency control using the Kharitonov's theorem," International Journal of Electrical Power & Energy Systems, vol. 55, pp. 51-58, 2014.

[9] S. Sondhi and Y. V. Hote, "Fractional order PID controller for perturbed load frequency control using Kharitonov's theorem," International Journal of Electrical Power & Energy Systems, vol. 78, pp. 884-896, 2016.

[10] M. Ayman and M. Soliman, "Robust multi-objective PSSs design via complex Kharitonov's theorem," European Journal of Control, vol. 58, pp. 131-142, 2021.

[11] K. Sharma, A. K. Yadav, and B. B. Sharma, "Kharitonov theorem-based robust control approach for sustainable microgrid against DoS cyber-attack," Digital Chemical Engineering, 2023, 7.

[12] N. M. Dehkordi, N. Sadati, and M. Hamzeh, "Robust tuning of transient droop gains based on Kharitonov's stability theorem in droop-controlled microgrids," IET Generation, Transmission & Distribution, vol. 12, no. 14, pp. 3495-3501, 2018.

[13] N. Ravikumar and G. Saraswathi, "Robust Controller Design for Speed Regulation of a Wind Turbine using 16-Plant Theorem Approach," EAI Endorsed Transactions on Energy Web, vol. 6, no. 24, 2019.

[14] S. R. Potturu and R. Prasad, "Model Order Reduction of LTI Interval Systems Using Differentiation Method Based on Kharitonov's Theorem," IETE Journal of Research, vol. 68, no. 3, pp. 2079-2095, 2019.

APPENDIX

Parameters related to (2):

$$A = L_l^2 L_{f1} + 2L_l L_{f1} L_{f2} + L_{f2}^2 L_{f1}$$

$$B = L_{f2}^2 R_{f1} + 2R_{f1} L_l L_{f2} + 2R_{f2} L_{f1} L_{f2} + 2R_l L_{f1} L_{f2} + R_{f2} L_{f1} L_l + R_l L_{f1} L_l + L_l^2 R_{f1} + L_l L_{f1} L_{f2} + L_{f2}^2 R_c + 2L_{f2} L_l R_c + L_l^2 R_c + R_c L_{f1} L_{f2} + R_c L_{f1} L_l$$

$$C = 2L_{f2} R_{f1} R_{f2} + 2L_{f2} R_{f1} R_l + 2L_l R_{f1} R_{f2} + R_{f2}^2 L_{f1} + 2L_{f1} R_l R_{f2} + L_{f1} L_l R_l + R_l^2 L_{f1} + L_l R_l R_{f1} + L_{f2}^2 / C_f + 2L_{f2} L_l / C_f + L_l^2 / C_f + 2L_{f2} R_c R_{f2} + 2L_{f2} R_l R_c + 2L_l R_c R_{f2} + 2L_l R_l R_c + R_{f1} R_c L_{f2} + R_{f1} R_c L_l + L_{f1} L_{f2} / C_f + L_{f1} L_l / C_f + L_{f1} R_c R_{f2} + L_{f1} R_c R_l$$

$$H = R_{f1} L_{f2} + R_{f1} L_l / C_f + R_{f1} R_c R_{f2} + R_{f1} R_c R_l + L_{f1} R_{f2} + L_{f1} R_l / C_f + 2L_{f2} / C_f + 2L_{f2} R_l / C_f + 2R_{f2} L_l / C_f + 2L_l R_l / C_f + R_{f2}^2 R_c + 2R_{f2} R_l R_c + R_l^2 R_c - R_c^3 + R_{f2}^2 R_{f1} + 2R_{f2} R_{f1} R_l + R_{f1} R_l^2$$

$$F = R_{f1} R_{f2} + R_{f1} R_l / C_f - 3R_c^2 / C_f + R_{f2}^2 / C_f + 2R_{f2} R_l / C_f + R_l^2 / C_f$$

$$E = 3R_c / C_f^2$$

$$D^* = 1 / C_f^3$$

$$K = L_{f2}^2 R_c + R_c L_{f2} L_l + L_l L_{f2} R_c + L_l^2 R_c$$

$$L = L_{f2}^2 / C_f + L_{f2} L_l / C_f + L_{f2} L_l / C_f + L_l^2 / C_f + R_c L_{f2} R_{f2} + R_c L_{f2} R_l + R_c L_{f2} R_{f2} + R_c L_l R_{f2} + R_l L_{f2} R_c + R_l L_l R_c + R_c L_l R_{f2} + R_c L_l R_l + R_c^2 L_{f2} + R_c^2 L_l$$

$$O = L_{f2} R_{f2} / C_f + L_{f2} R_l / C_f + L_{f2} R_{f2} / C_f + R_{f2} L_l / C_f + L_{f2} R_l / C_f + L_l R_l / C_f + L_l R_{f2} / C_f + L_l R_l / C_f + R_c R_{f2}^2 + R_{f2} R_l R_c + R_{f2} R_l R_c + R_c R_l^2 + 2L_{f2} R_c / C_f + R_{f2} R_c^2 + R_l R_c^2 + 2R_c L_l / C_f$$

$$N = L_{f2} / C_f^2 + L_l / C_f^2 + 2R_{f2} R_c / C_f + 2R_l R_c / C_f + R_{f2}^2 / C_f + R_{f2} R_l / C_f + R_l R_{f2} / C_f + R_l^2 / C_f$$

$$M = R_{f2} / C_f^2 + R_l / C_f^2 .$$

Parameters related to (4):

$$Y = K_{i1} K_{i2} J - K_{i1} K_{i2} J w_g$$
$$Z = J^2 w_g$$
$$K = K_{i1} D J w_g .$$

Parameters related to (5):

$$a_6 = J^2 L_{f2}^2 R_c + R_c L_{f2} L_l + L_l L_{f2} R_c + L_l^2 R_c w_g + K_{i2} J L_l^2 L_{f1} + 2L_l L_{f1} L_{f2} + L_{f2}^2 L_{f1} K_{i1} - K_{i1} w_g$$

$$a_5 = J^2 L_{f2}^2 / C_f + L_{f2} L_l / C_f + L_{f2} L_l / C_f + L_l^2 / C_f + R_c L_{f2} R_{f2} + R_c L_{f2} R_l + R_c L_{f2} R_{f2} + R_c L_l R_{f2} + R_l L_{f2} R_c + R_l L_l R_c + R_c L_l R_{f2} + R_c L_l R_l + R_c^2 L_{f2} + R_c^2 L_l w_g + D J L_{f2}^2 R_c + R_c L_{f2} L_l + L_l L_{f2} R_c + L_l^2 R_c K_{i1} w_g + K_{i2} J L_{f2}^2 R_{f1} + 2R_{f1} L_l L_{f2} + 2R_{f2} L_{f1} L_{f2} + 2R_l L_{f1} L_{f2} + R_{f2} L_{f1} L_l + R_l L_{f1} L_l + L_l^2 R_{f1} + L_l L_{f1} L_{f2} + L_{f2}^2 R_c + 2L_{f2} L_l R_c + L_l^2 R_c + R_c L_{f1} L_{f2} + R_c L_{f1} L_l K_{i1} - K_{i1} w_g$$

$$a_4 = J^2 L_{f2} R_{f2} / C_f + L_{f2} R_l / C_f + L_{f2} R_{f2} / C_f + R_{f2} L_l / C_f + L_{f2} R_l / C_f + L_l R_l / C_f + L_l R_{f2} / C_f + L_l R_l / C_f + R_c R_{f2}^2 + R_{f2} R_l R_c + R_{f2} R_l R_c + R_c R_l^2 + 2L_{f2} R_c / C_f + R_{f2} R_c^2 + R_l R_c^2 + 2R_c L_l / C_f w_g + D J L_{f2}^2 / C_f + L_{f2} L_l / C_f + L_{f2} L_l / C_f + L_l^2 / C_f + R_c L_{f2} R_{f2} + R_c L_{f2} R_l + R_c L_{f2} R_{f2} + R_c L_l R_{f2} + R_l L_{f2} R_c + R_l L_l R_c + R_c L_l R_{f2} + R_c L_l R_l + R_c^2 L_{f2} + R_c^2 L_l K_{i1} w_g + K_{i2} J 2L_{f2} R_{f1} R_{f2} + 2L_{f2} R_{f1} R_l + 2L_l R_{f1} R_{f2} + R_{f2}^2 L_{f1} + 2L_{f1} R_l R_{f2} + L_{f1} L_l R_l + R_l^2 L f 1 + L_l R_l R_{f1} + L_{f2}^2 / C_f + 2L_{f2} L_l / C_f + L_l^2 / C_f + 2L_{f2} R_c R_{f2} + 2L_{f2} R_l R_c + 2L_l R_c R_{f2} + 2L_l R_l R_c + R_{f1} R_c L_{f2} + R_{f1} R_c L_l + L_{f1} L_{f2} / C_f + L_{f1} L_l / C_f + L_{f1} R_c R_{f2} + L_{f1} R_c R_l K_{i1} - K_{i1} w_g$$

$$a_3 = J^2 L_{f2} / C_f^2 + L_l C_f^2 + 2R_{f2} R_c / C_f + 2R_l R_c / C_f + R_{f2}^2 / C_f + R_{f2} R_l / C_f + R_l R_{f2} / C_f + R_l^2 / C_f w_g + D J L_{f2} R_{f2} / C_f + L_{f2} R_l / C_f + L_{f2} R_{f2} / C_f + R_{f2} L_l / C_f + L_{f2} R_l / C_f + L_l R_l / C_f + L_l R_{f2} / C_f + L_l R_l / C_f + R_c R_{f2}^2 + R_{f2} R_l R_c + R_{f2} R_l R_c + R_c R_l^2 + 2L_{f2} R_c / C_f + R_{f2} R_c^2 + R_l R_c^2 + 2R_c L_l / C_f K_{i1} w_g + K_{i2} J R_{f1} L_{f2} + R_{f1} L_l / C_f + R_{f1} R_c R_{f2} + R_{f1} R_c R_l + L_{f1} R_{f2} + L_{f1} R_l / C_f + 2L_{f2} / C_f + 2L_{f2} R_l / C_f + 2R_{f2} L_l / C_f + 2L_l R_l / C_f + R_{f2}^2 R_c + 2R_{f2} R_l R_c + R_l^2 R_c - R_c^3 + R_{f2}^2 R_{f1} + 2R_{f2} R_{f1} R_l + R_{f1} R_l^2 K_{i1} - K_{i1} w_g$$

$$a_2 = J^2 R_{f2} / C_f^2 + R_l / C_f^2 w_g + D J L_{f2} / C_f^2 + L_l / C_f^2 + 2R_{f2} R_c / C_f + 2R_l R_c / C_f + R_{f2}^2 / C_f + R_{f2} R_l / C_f + R_l R_{f2} / C_f + R_l^2 / C_f K_{i1} w_g + K_{i2} J R_{f1} R_{f2} + R_{f1} R_l / C_f - 3R_c^2 / C_f + R_{f2}^2 / C_f + 2R_{f2} R_l / C_f + R_l^2 / C_f K_{i1} - K_{i1} w_g$$

$$a_1 = D J R_{f2} / C_f^2 + R_l / C_f^2 K_{i1} w_g + K_{i2} J 3R_c / C_f^2 K_{i1} - K_{i1} w_g$$

$$a_0 = -1 / C_f^3 .$$

Creation of All-Season Photothermal Installation of Increased Efficiency

Muhammadjon Tursunov[1], Habibullo Sabirov[1], Tohir Axtamov[1], Mamasobir Tursunov[2], Maxamadi Chariyev[2] and Habib Abdullayev[3]
[1]*Physical-Technical Institute, Chingiz Aytmatov Str. 2B, Tashkent, Uzbekistan*
[2]*Termez State University, Barkamol Avlod Str. 43, Termez, Uzbekistan*
[3]*Chirchik State Pedagogical University, Amir Temur Str. 104, Chirchik, Uzbekistan*
muhammadtursunov54@gmail.com, sabirovhabibullo47@gmail.com, tohiraxtamov@gmail.com, mtursunov@tersu.uz, chariyevmaxamadi@gmail.com, habibabdullayev04@gmail.com

Keywords: Photovoltaic, Photovoltaic Systems, Photo Thermal, Monocrystalline, Heat Collector, Reflectors, Mobile Portable Device, Solar Radiation, Short Circuit Current, Open Circuit Voltage.

Abstract: In this article, a project of portable photoelectric and photo thermal devices with a power of 60W based on crystalline silicon photoelectric batteries for the production and use of electricity and hot water for household needs in the rural areas of the republic was created. This natural experiment served two purposes. The first is to determine the optimal type for the climate when the various types of PVs available in the local commercial market are used in hot climates. Second, the results of determining and comparing the changes in parameters and energy efficiency of PV and PVT with the same base and capacity at the same time under the same conditions are presented. The results of the study of the parameters that increase their effectiveness have been considered. Special attention is paid to modern components and equipment so that these devices are relatively compact and convenient to operate and control occurs easily.

1 INTRODUCTION

The efficiency of using photovoltaic systems depends on the efficiency of using photovoltaic arrays, energy storage systems, control and monitoring electronics, and the climatic conditions of the area. In rural regions of Uzbekistan, where the issue of supplying the population and farms with electricity and water for household and domestic needs is especially acute, the efficiency of using photovoltaic installations is largely determined by climatic conditions [1]. It is expressed in sharp amplitudes of day and night, summer and winter temperatures. The nature is arid, there is little precipitation, low relative humidity. The duration of a sunny day in summer (from the moment the Sun appears above the horizon till the visible sunset) is more than 13 hours, in winter - at least six hours. The coldest month is January, when the temperature in the north drops to 10 or more degrees below zero, and in the extreme south, near the city of Termez, it is above zero. The absolute minimum of winter temperatures is 35-38 degrees below zero. The hottest month is July, and in the mountainous regions is July-August. The average temperature during this period on the plains and foothills is 25-30 degrees above zero, and in the south (Termez and Sherabad) it reaches 41-42 degrees above zero [2]. From a comparison of sunshine during the daytime in summer and winter, it can be concluded that in summer the energy generated by photovoltaic arrays (PVA) is more than twice the energy in winter. In addition, the intensity of solar radiation in summer is twice the intensity of radiation in winter, due to the angle of incidence of radiation on the surface of the PVA.

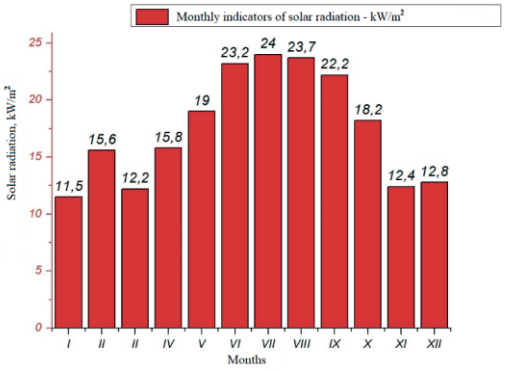


Figure 1: Monthly changes of the average intensity of solar radiation along the normal to the surface for Tashkent.

Figure 1 shows the monthly changes of the average intensity of solar radiation along the normal to the surface for Tashkent [3]. The deviation for the month of March is due to the relatively numerous cloudy and rainy days in this month (Figure 2). The trend of the distribution of temperature and precipitation by months for the overwhelming rural regions of Uzbekistan (to the west area of the country) will be similar to the trend for the city of Tashkent (Figure 2). This is due to the fact that the wind flow is directed from the west (Karakalpakstan) to the east towards Tashkent.

It can be concluded that from November to March we have low temperatures, which is favorable for the operation of the PVA, but at the same time, the prevailing cloudy weather is unacceptable for the efficient operation of the PVA. In addition, this time is characterized by relatively short days and low intensities of solar radiation. The southern region also has some features of the continental climate, such as a large difference in night and day temperatures, a large number of cloudy days and frequent fogs in the winter season. In general, these factors hinder the efficiency of traditional photovoltaic installations and their overall efficiency. Monthly, on average, it decreases by 50% or more in the winter months, and by more than 20-30% in the autumn-spring months [4]. And in the summer, the difference increases to 30-40%. In this regard, there is a need to create a new, more advanced energy supply system for rural regions based on photovoltaic installations equipped with photothermal batteries and reflectors of various designs.

2 METHODS AND MATERIALS

This paper presents the results of research and development of autonomous mobile photothermal installation of increased efficiency for use in the dry climate of Uzbekistan for all seasons of the rural regions of the republic.

In rural regions of the country there are specific factors that differ from the conditions of use the installations in other countries. Such factors are a large temperature difference between day and night, the number of cloudy days and, accordingly, precipitation in the form of rain and snow, and the presence of seasonal dust storms. In such conditions, supplying electricity and water to rural households and creating the necessary comfort for residents is an undeniably difficult task. From Figure 1 and Figure 2, it can be concluded that 5 months of the year (November-March) are unfavorable in terms of precipitation and solar radiation intensity for the use of photovoltaic (PVB), photothermal batteries (PTB) and installations based on them. From Figure 1 it can be seen that the minimum monthly indicators of solar radiation fall on the cold and cloudy and relatively short November-March months of the year. Although the temperature indicators are comfortable for PVB operation, however, the number of rainy and cloudy days is about half of the time period. Daylight hours are 2 times less in duration compared to the summer period of the year [5].

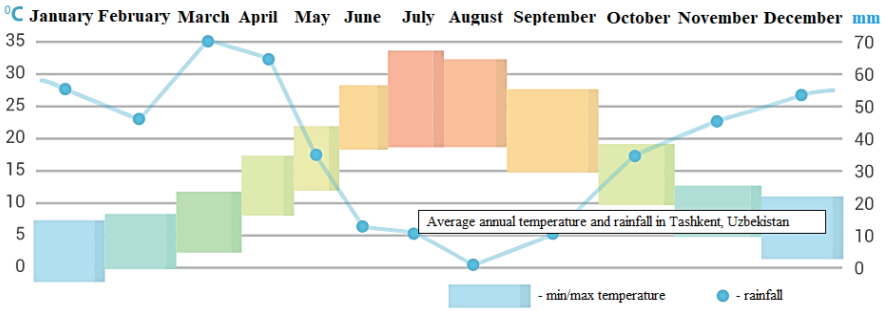


Figure 2: Distribution of temperature and precipitation by months for the conditions of Tashkent.

To determine the required level of solar radiation intensity for generating electrical energy and hot water in the winter months of the year, it is necessary to determine the applicability limit of frontal mesh contacts of solar cells (SC) of PVB based on crystalline silicon with an increase in the intensity of radiation incident on the surface of the battery. According to the condition AM 1, the short-circuit current of the SC is measured at 1000W/m^2 and a temperature of 25^{0}C. Grid contacts of modern solar cells made of crystalline silicon ensure the linearity of the short-circuit current up to the specified value of the illumination level of the AM 1 condition.

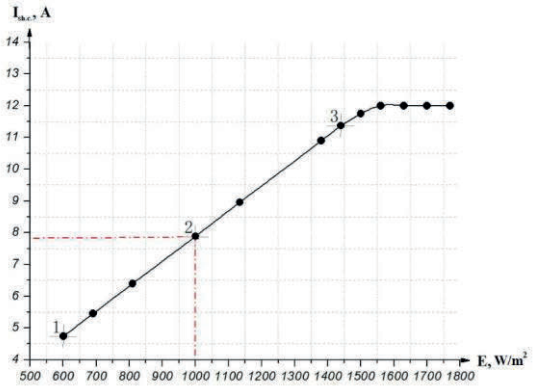


Figure 3: Dependence of the SC short-circuit current on the intensity of solar radiation. 1-Measurement start at 600 W/m^2 solar intensity, 2-Short circuit current at 1000 W/m^2, 3-Short circuit current bending point.

We have studied the dependences of the short-circuit current on the intensity of solar radiation by 40-50% more than in the AM 1 condition incident on the surface of a solar cell made of single-crystal silicon, which is shown in Figure 3. The experiments were carried out in natural conditions using sunlight. Solar cells sized 156x156mm^2 are installed on a cooled metal plate, to increase (change) the intensity of solar radiation, reflectors with a solar radiation reflection coefficient of 0.5 are used. As can be seen from Figure 3, the dependence of the SC short-circuit current on the intensity of solar radiation depends linearly up to a radiation intensity of 1450–1500W/m^2 without additional influence on the electrical properties of the front current-collecting contacts [6]. Thus, in the winter months of the year, with an increase in the intensity of solar radiation to the indicated values at a temperature of SC 25^{0}C during short winter daylight hours in clear sunny weather for 5-6 hours, with an efficiency of 19-20%, it is possible to obtain up to 1500W/m^{2}x1m^{2}x0.19x5h=1425W per day of energy from 1 m^2 of the PVB area, which is approximately equal to the energy received from the PVB with an area of 1m^2 in the summer at radiation intensity ~ 800W/m^2 for 8-9 hours. An increase in the value of the intensity of solar radiation to more than 1500W/m^2 leads to a loss of short-circuit current, respectively, power due to the non-linearity of the characteristic of the dependence of the short-circuit current on the intensity of solar radiation.

2.1 Installation Options

To conduct research and develop a photothermal installation with increased efficiency in the winter months of the year, there was used a low-power portable autonomous photovoltaic installation with a power of 60W [2-3]. The area of the photovoltaic battery is 0.36m^2. Four additional reflectors (two side ones, top and bottom) with a total area equal to two areas of the photovoltaic battery (Figure 4) were made on the chosen design of the installation to change and select the optimal value of the intensity of solar radiation incident on the surface of the photovoltaic battery. The PVB temperature was controlled by changing the water supply rate of the thermal collector installed on the rear surface of the photovoltaic battery.

a b

Figure 4: Four additional reflectors: a) Portable photovoltaic installation (initial), b) portable photothermal installation for the winter period of the year.

In order to determine the number of reflectors used depending on the months of the year, there were studied the dependence of solar radiation intensity on time by months of the year. For example, for the winter season, there was studied the dependence of solar radiation intensity on time on November 14-15th, 2022 (Figure 5.). There was determined the number of reflectors by months of the year using the data. In the winter season (November-March), the number of reflectors used

should be four, and in the spring (April-May) and autumn (September-October) the number of reflectors should be two, while the upper and lower reflectors are removed, which are removable. We consider it optimal to install two side reflectors in the structure in the winter season. During daylight hours, the Sun is on the south side and therefore orientation in the direction of the Sun will be easier. For example, one of the options is to install two side reflectors at an angle of 120^0 to each other, then the side reflectors will be directed towards the Sun during daylight hours. The photo thermal battery is made according to the technology given in [7], cellular polycarbonate with parallel channels is used as a thermal collector. In the summer months, this PTB design is also used, which, when using two reflectors, provides electrical energy that is 1.5-1.6 times greater than the use of PTBs without reflectors [8]. Efficient heat removal and the possibility of obtaining, if necessary, an outlet water temperature of up to 60^0C; when using two side reflectors, the output of water can be increased up to 20liters/hour.

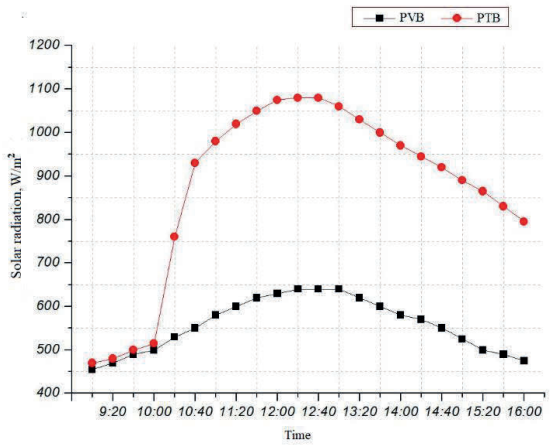


Figure 5: The dependence of solar radiation intensity on time.

3 RESULTS AND DISCUSSION

Figure 5 shows the dependence of solar radiation intensity on time dated on November 14th, 2022 (direct solar radiation), falling on the surface of the PVB in the interval from 9:00 to 16:00 400W/m^2 during the day. The maximum of solar radiation intensity 600-650W/m^2 becomes in the time interval from 11:20 to 13:30. When connecting 4 reflectors, the maximum value of solar radiation intensity reaches approximately 1100W/m^2. This value corresponds to the solar radiation intensity of photothermal batteries with two reflectors in the summer. Dependences of solar radiation intensities for direct solar radiation and enhanced with the use of four reflectors have a similar distribution over time, except for the interval from 10:00 to 10:40 during the deployment of reflectors.

Figure 6 shows the dependences of open-circuit voltages and short-circuit currents for PVBs and PTBs based on PVBs on time. As can be seen from Figure 6, the dependences of the short-circuit currents of the PVB and PTB are similar to the dependences of solar radiation intensity (Figure 5). The exception is the dependence of the open-circuit voltage of the PTB, starting from the time of deployment of four reflectors, leading to a sharp decrease in the voltage value from 21.65V to 17.3V due to heating of the rear surface of the PVB. Later, when cold water is connected to the heat collector, the open-circuit voltage increases to a value of 20.8V, as in. Water temperature can be changed within 25-50^0C. The dependence of the short-circuit current on time repeats the dependence of solar radiation intensity on time. The short circuit current of the PTB in the time interval from 11:30 to 13:00 is twice the short circuit current of the photovoltaic battery.

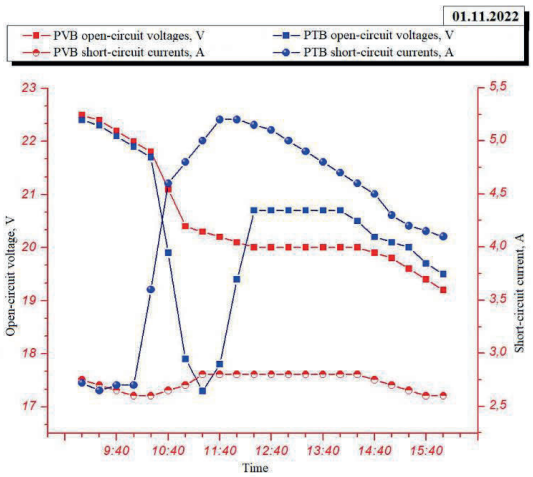


Figure 6: Dependences of open-circuit voltages and short-circuit currents for PVB and PTB based on PVB on time.

Figure 7 shows the dependence of the generated electric power on time for PVB and PTB. As can be seen, for the case of PVB, the dependence of power is largely determined by the dependence of the short-circuit current on time (Figure 6). In general, the electrical power of the PVB, generated from 10:00 to 14:40, is equal to approximately 40W per hour. For the PTB until 10 am, the power generated is the same as the power generated by the PVB.

Further, when opening 2 reflectors (upper and lower), the power increases to 63W (~ 1.5 times), and a further decrease in power is associated with heating of the PVB, which leads to a decrease in the open circuit voltage.

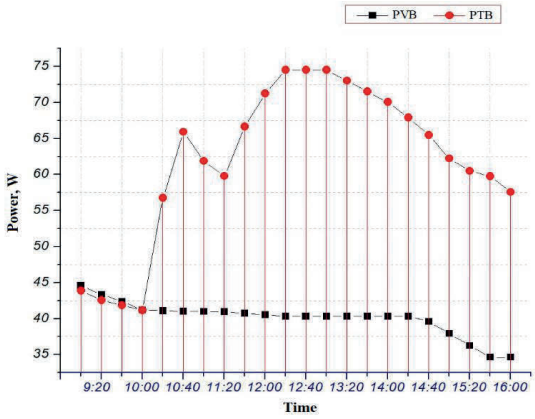


Figure 7: Dependence of the power for PVB and PTB on time.

When water is connected to the heat collector and two more side reflectors are turned on, the PTB power increases to 76W. After 13 hours and 30 minutes, the electrical power decreases due to a decrease in the intensity of solar radiation (with the appearance of scattered white "spots" in the sky). Generally, in clear weather in the winter season, the developed installation based on the PTB makes it possible to obtain an electric power that is almost twice the power of the PVB [9]. At the same time, if necessary, it is possible to obtain heated water up to a temperature of 40-50^{0}C, which is important for the population of rural regions [10]. Therefore, the development and use of such photo thermal installations with a power of up to 1000W is in great demand for rural residents of the regions of the country. In the future, there will be studied the efficiency of using such installations of low intensity of solar radiation and temperatures, which, according to Figure 1 and Figure 2 fall on January-April and October-December months of the year for rural regions of the country.

4 CONCLUSIONS

There has been made an analysis of the influence of climatic conditions in the regions of Uzbekistan on the possibility of using photovoltaic installations in all seasons of the year. For this purpose, an autonomous mobile photo thermal installation has been developed for use in the dry climate of Uzbekistan for all seasons of the country's rural areas. It was revealed that the efficiency of use (generated electric power) of the developed photo thermal installation in the winter period of the year (November) is two times higher than the electric power of a similar photovoltaic installation. The photo thermal installation, if necessary, also produces hot water with a temperature of up to 50^{0}C.

REFERENCES

[1] R.R. Avezov, N.R. Avezova, and E.Yu. Rakhimov, "Taking into account the orientation of heated rooms in the calculation of insolation passive heating systems," International Conference "Fundamental and Applied Problems of Physics", Tashkent, June 13-14, 2017, pp. 5-10.

[2] M.N. Tursunov, Kh. Sabirov, T.Z. Akhtamov, and B.A. Yuldoshev, Patent: Portable low power photovoltaic device. SAP 2021 0236.

[3] R.A. Muminov, M.N. Tursunov, Kh. Sabirov, T.Z. Akhtamov, and M. Eshmatov, "Comprehensive improvement of the efficiency of a mobile photovoltaic installation for water lifting through the use of photothermal batteries, side reflectors of solar radiation, and cooling water from deep underground aquifers," ISSN 0003-701X, Applied Solar Energy, vol. 58, no. 2, 2022, pp. 238-243.

[4] R.Yu. Akbarov and A.A. Kuchkarov, "Modeling and calculation of optical-geometric characteristics of a solar concentrator with flat fresnel mirrors," Applied Solar Energy, 54(3). 2018, pp.187-192.

[5] L.T. Kostic, T.M. Pavlovic, and Z.T. Pavlovic, "Influence of Reflectance from Flat Aluminum Concentrators on Energy Efficiency of PV/Thermal Collector," Appled Energy, 2010, vol. 87, pp. 410-416.

[5] V.P. Anand, Mohd. and Meraj Khan, et. al, "Performance Improvement of Solar Module System using Flat Plate Relectors," ICAEE, 2014.

[6] R. Arshad, S. Tariq, M.U. Niaz, and M. Jamil, "Improvement in solar panel efficiency using solar concentration by simple mirrors and by cooling," iCREATE, 2014.

[7] D. Augustin, R. Chacko, and J. Jacob, "Canal Top Solar PV with Reflectors," IEEE International Conference PEDES, 2016.

[8] N.V. Revzina, "Development of algorithms for calculating the optimal geometry of reflectors," Omsk region – a deposit of opportunities. Mat. V regional youth scientific and practical. conf. Omsk, 2014.

[9] L.T. Kostic, T.M. Pavlovic, and Z.T. Pavlovic, "Optimal Design of Orientation of PV/T Collector with Reflectors," Applied Energy, 2010, pp. 87. 3023-3029.

[10] D.T.P. Wijesuriya, K.D.S.H. Wickramathilaka, L.S. Wijesinghe, D.M. Vithana, and H.Y. Ranjit Perera, "Placeing Reflectors for Reducing Payback Period of Solar PV for Smart Buildings," IEEE 15th INDIN, 2017.

Optimal Parameter Determination Asynchronous Traction Engine to Improve Operating Performance

Nurali Pirmatov[1,3], Komil Usmonov[2], Usan Berdiyev[2], Tulagan Nazirkhonov[2] and Ulmasbek Berdiyorov[2]
[1]*Tashkent State Technical University, Universitet Str. 2, Tashkent, Uzbekistan*
[2]*Tashkent State Transport University, Temiryolchilar Str. 1, Tashkent, Uzbekistan*
[3]*University of Tashkent for Applied Sciences, Gavhar Str. 1, Tashkent, Uzbekistan*
nurali.pirmatov@mail.ru, nazirkhonov.t@gmail.com, komil.usmonov94@gmail.com

Keywords: Electric Locomotive, Asynchronous Traction Engine, Electric Power, Electric Locomotive "O'zbekiston Yo'lovchi".

Abstract: Increasing the energy performance of rolling stock is one of the most relevant railway transport. One of the ways to achieve these tasks is to determine the optimal parameters of the traction drive of the rolling stock, with which you can create computer simulation models. In this article, the optimal parameters of an asynchronous traction motor of an electric rolling stock are determined. Determination of the parameters of the ATM is necessary to create a computer simulation model that allows reproducing electromagnetic processes in a traction electric drive and converters, as well as processing functions of the simulation results obtained that are adequate to the real conditions of use on electric rolling stock of converters with various control algorithms in traction and regenerative braking modes. In the article, the modes of idling, short circuit and rated load are used to determine the parameters of the ATM in relation to the T-shaped replacement circuit using the method of separation of losses in the engine. Losses in traction converters and traction gearboxes are taken into account in accordance with the power developed by the ATM. The results obtained can be used in a computer simulation model designed to reproduce electromagnetic processes in a traction electric drive and converters, when determining the energy characteristics of electric locomotives with asynchronous traction motors.

1 INTRODUCTION

In the years after independence, the construction of new and electrification of existing railways has been developing rapidly in the Republic of Uzbekistan [1-5]. To date, the whole republic is permeated with threads of railways that make up about 6.15 thousand km, of which 1,783 thousand km have been electrified. To replace the outdated electric locomotives of the “VL” series with collector traction motors, electric locomotives of the “O'zbekiston” series produced by “P.R.China” with asynchronous traction motors (ATM) were brought. The new family of electric locomotives are characterized by greater power and high efficiency due to ATM, which are controlled by “GTO” thyristors for the first generation of electric locomotives and “IGBT” transistors for subsequent ones. For passenger high-speed traffic up to 160 km/h, electric locomotives “O'zbekiston Yo'lovchi” are operated, shown in Figure 1.

One of the urgent problems of railway transport is to increase the energy performance of rolling stock. To achieve these goals, it is necessary to determine the characteristics of the traction drive, as well as to create a computer model to simulate electromagnetic processes in converters and electric drives. Losses and expenses of electric locomotive electricity can be determined only by analytical methods.

Figure 1: Electric locomotive "O'zbekiston Yo'lovchi".

2 METHODS AND MATERIALS

The determination of the parameters of the ATM windings is mostly complicated by the technical conditions of the test modes. Based on the no-load and short-circuit modes, by the method of separation of losses in the rated load mode, it is possible to determine analytically the parameters of the ATM active resistance, the inductance of scattering of the stator and rotor windings, losses in steel, in stator copper and mechanical losses.

The method of analytical calculation was used with further development of a model of a traction asynchronous motor in the MATLAB Simulink package. The resistance of the stator phase windings in this case is determined by the loss power ΔP_{smn} in copper at rated load [6-10], (ohms)

$$r_s = \frac{\Delta P_{SM.N}}{3I_{SN}^2}, \quad (1)$$

where I_{sn}- is the current of the stator phase in the nominal mode.

We determine the power loss in the copper of the stator at rated load by the formula (W)

$$\Delta P_{sm.n} = \frac{P_n \times 10^3}{\eta_n}\left(1 - (\eta_n + s_n)\right) - (\Delta P_{mech.n} + \Delta P_{s.ct.n} + \Delta P_{s.dm}). \quad (2)$$

where s_n is the sliding of the rotor in the rated load mode, P_n is the power of the rated mode, $\eta_н$ is the efficiency of the rated mode, $\Delta P_{s.ct.n}$ is the power of losses in the stator steel in the rated load mode, $\Delta P_{mech.n}$ is the power of mechanical losses in the rated load mode, $\Delta P_{s.dm}$ is the power of additional losses in the stator copper due to the spatial harmonics of the stator current.

The power of mechanical losses in the rated load mode can be taken $\Delta P_{mech.n} = 0{,}002\frac{P_n}{\eta_n} \times 10^3$ (W), the power of additional losses from spatial harmonics of the current $\Delta P_{s.dm} = 0{,}005\frac{P_N}{\eta_N} \times 10^3$W.

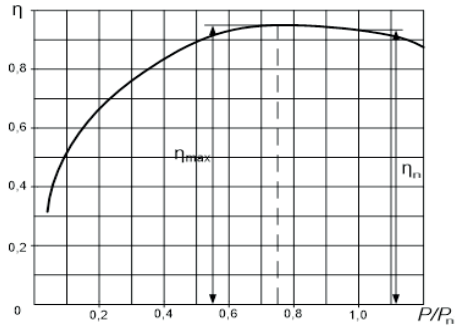


Figure 2: Dependence of the efficiency of the ATM on the load.

The power of losses in steel in the rated load mode is almost equal to the power of losses in steel for the rated load mode [11]. The latter can be determined from the condition of equality of the main losses in steel and copper of the stator at the maximum efficiency of the motor n_{max} corresponding to the mode of the rated load [12-16] (Figure.2), (W):

$$\Delta P_{s.ct.\frac{3}{4}} = \frac{3}{4 \times 2}\frac{P_n \times 10^3}{\eta_{max}}(1 - \eta_{max}) - (\Delta P_{mech.\frac{3}{4}} + \Delta P_{sdm}). \quad (3)$$

Mechanical losses at the nominal rotor speed practically do not depend on the load, therefore, it can be assumed with sufficient accuracy that $\Delta P_{mech.3/4} = \Delta P_{mech.n}$.

The resistance of the rotor, reduced to the resistance of the stator phase, can be calculated by the power of losses in the rotor at nominal slip and the reduced rotor current I'_{rn} (ohms):

$$r'_r = \frac{P_n \times 10^3 \times s_n}{3I'^2_{rn}}. \quad (4)$$

We find the rotor current reduced to the stator current for the rated load mode according to the formula (A):

$$I'_{rn} = I_{sn}\sqrt{1 - \left(\frac{I_{s0}}{I_{sn}}\right)^2} - \frac{I_{s0}}{I_{sn}2b_n}. \quad (5)$$

In which $I_{s0}$is the idling current of the ATM at the rated voltage and the rated frequency of the stator current, b_n is the multiplicity of the greatest electromagnetic moment.

The value of $b_n$is determined by the rated load mode and the inductive short-circuit resistance $X_k$of the motor:

$$b_n = \frac{U_{sn}^2(1-s_n)}{\sqrt{3} \times 2c_1(P_n \times 10^3 + \Delta P_{mech.n})X_k}, \quad (6)$$

where c_1 is the coefficient of reduction of the parameters of the ATM windings to the G-shaped substitution scheme.

The total short-circuit resistance for the ATM stator phase is determined by the multiplicity of the short-circuit current $K_{1s/c}$=7-8 (ohms), equal to:

$$Z_k = \sqrt{X_k^2 + (c_1 r_k)^2} = \frac{U_{sn}}{\sqrt{3}I_{sn}K_{1s/c}}, \quad (7)$$

here r_k is the short circuit resistance of the motor, r_k=r_s+r_r'.

The inductive resistance of a short circuit can be determined from the formula (ohms):

$$X_k = Z_k\sqrt{1 - cos^2\varphi_{s/c}}. \quad (8)$$

In which $cos\ \varphi_{s/c}$ is the motor power factor for the short circuit mode, we preliminarily accept $cos\ \varphi_{s/c} = 0{,}15 \sim 0{,}2$.

Idle current ATM (A):

$$I_{s0} \approx I_{sn}(sin\varphi_{sn} - \frac{cos\varphi_{sn}}{2b_n}), \quad (9)$$

where $\cos\varphi_{sn}$ is the engine power factor for the rated mode.

After performing calculations according to (1) – (9), the values of $r_r', X_k, c_1 \approx 1 + \frac{Z_k}{2Z_0}$ should be clarified and the refined calculation of the active and inductive resistances of the stator and rotor of the ATM should be repeated.

From the experience of designing an ATM with a capacity of 1-1.3 MW, the ratio of the stator phase scattering inductance and the reduced rotor phase scattering inductance in relative units is 1: 0.8.

Stator phase scattering inductance ($L_{s\sigma}$ (H)):

$$L_{s\sigma} = \frac{X_k}{1{,}8 \times 2\pi \times f_{sn}}, \quad (10)$$

where f_{sn} is the frequency of the stator voltage in the nominal mode.

The rotor phase scattering inductance ($L'_{r\sigma}$(H)) reduced to the stator phase is equal to:

$$L'_{r\sigma} = \frac{X_k \times 0{,}8}{1{,}8 \times 2\pi \times f_{sn}}. \quad (11)$$

The total resistance of the motor phase in idle mode (Z_0, ohms) :

$$Z_0 = \frac{U_{sn}}{\sqrt{3} \times I_{s0}}. \quad (12)$$

Active component of the no-load current of the stator phase, (I_{s0a}, (A)) :

$$I_{s0a} = \frac{\Delta P_{s.ct.3/4} + (r_s \times I_{s0}^2) + \Delta P_{s.dm} + \Delta P_{mech.n}}{\sqrt{3} \times U_{sn}}. \quad (13)$$

Magnetization current of the stator phase, (I_μ, A)):

$$I_\mu = \sqrt{I_{s0}^2 - I_{s0a}^2}. \quad (14)$$

The mutual induction resistance (X_μ, ohms) of the stator and rotor phase windings is calculated by the (15):

$$X_\mu = \frac{U_{sn}}{\sqrt{3} \times I_\mu} - X_{s\sigma}. \quad (15)$$

In which $X_{s\sigma}$ is the inductance resistance of the stator phase scattering, $X_{s\sigma} = 2\pi f_{sn} L_{s\sigma}$.

The mutual inductance of the stator and rotor phase winding, reduced to the stator (L_μ, H) is equal to:

$$L_\mu = \frac{X_\mu}{2\pi f_{sn}}. \quad (16)$$

3 RESULTS AND DISCUSSION

Based on formulas (10) - (16), the calculation of the main parameters of the ATM "1TB2624–0GA02" used on electric locomotives "O'zbekiston Yo'lovchi" was performed. Passport data, design parameters and electrical values characterizing the operating modes of ATM "1TB2624–0GA02" are given in Table 1 and Table 2.

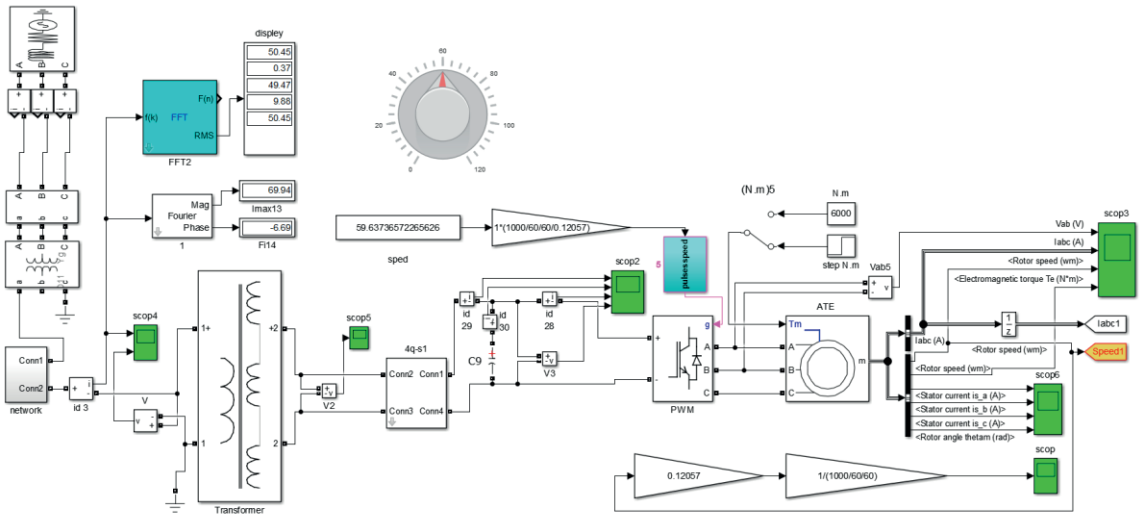


Figure 3: Computer model of an AC electric locomotive of a series "O'zbekiston Yo'lovchi" with 4q-S converters and automatic control system.

Table 1. Passport data of ATM "1TB2624-0GA02".

Parameters	Designation	Meaning
Rated power, kW	P_H	1020
Rotation speed, rpm	n_n	1484
Rated line voltage, V	U_{sn}	2063
Rated current, A	I_{sn}	330
Power factor at 100% load	$\cos \phi_{sn}$	0,85
Efficiency factor	η_n	0,95
Rated frequency of the stator current, Hz	f_{sn}	50
Number of pole pairs	p	2
Nominal rotor slip	s_n	0,011

Based on the data obtained, a mathematical model of the power supply and control scheme of the traction asynchronous motor "1TB2624-0GA02" of the electric locomotive "O'zbekiston Yo'lovchi" was created in the MATLAB Simulink software package, which is shown in Figure 3.

Simulation processes were performed at the voltage of the feeders of the traction substation of 27.5 kV and the greatest distance from the substation at 20 km, which corresponds to the equivalent electrical resistance of the network for a distance of 10 km with two-way power supply. The voltage of the DC link (output filters 4q-S of the converter) is taken equal to U_{dc}=2800 V. The load of the ATM was set by the value of the moment of resistance from $0.25M_{sn}$ to $1.25M_{sn}$ at a nominal frequency of the stator voltage of 50 Hz. Oscillograms of the stator voltage, stator currents, rotor speed, electromagnetic torque of ATM at rated load are shown in Figure 4.

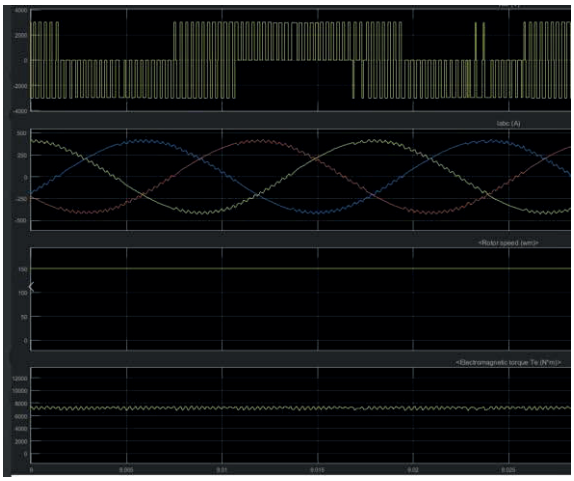


Figure 4: Oscillograms of stator voltage, stator currents, rotor speed, electromagnetic torque of ATM at rated load.

Table 2. Design parameters and basic electrical values of the ATM "1TB2624 – 0GA02".

Parameters and data	Designa-tion	Meaning
Loss power in the copper of the stator at rated load, W	$\Delta P_{sm.n}$	20132
The power of mechanical losses in the rated load mode, In	$\Delta P_{mech.n}$	2147
Power of additional losses from spatial harmonics of current, W	$\Delta P_{s.dm}$	5368
The power of losses in steel in the mode of ¾ of the rated load, W	$\Delta P_{s.ct.3/4}$	12616
The power loss in the rotor copper at rated load, W	$\Delta P_{rm.n}$	21742
Resistance of the stator phase windings, Ohms	r_s	0,059
The resistance of the rotor, reduced to the resistance of the stator phase, Ohms	r'_r	0,0337
Multiplicity of the greatest electromagnetic moment	b_n	1,9
Total short-circuit resistance, Ohms	Z_k	0,487
Inductive short-circuit resistance, Ohms	X_k	0,478
Inductive resistance of the stator phase scattering flows, Ohms	$X_{s\sigma}$	0,265
Inductive resistance of the rotor scattering flows reduced to the stator phase, Ohms	$X'_{r\sigma}$	0.212
Stator phase scattering inductance, H	$L_{s\sigma}$	0,0009181
The inductance of the scattering of the rotor phase, reduced to the stator phase, H	$L'_{r\sigma}$	0,0007345
The total resistance of the motor phase in idle mode, Ohms	Z_0	12,2017
Coefficient of reduction of winding parameters to the L-shaped substitution scheme	c_1	1,023
The rotor current reduced to the stator current for the rated load mode, A	I'_{rn}	349
Idle current, A	I_{s0}	99,12
The active component of the idle current of the stator phase, A	I_{s0a}	5,797
The magnetization current of the stator phase, A	I_μ	98,949
Inductive resistance of mutual induction of the stator and rotor phase windings, Ohms	X_μ	11,772
Mutual inductance of the stator and rotor, brought to the stator (magnetization inductance), H	L_μ	0,041

4 CONCLUSIONS

The study and development of ways to improve the energy efficiency of rolling stock is relevant. To increase the energy efficiency of rolling stock, it is necessary to determine the characteristics of the

traction drive, as well as to create a computer model for modeling electromagnetic processes in converters and electric drives. Losses and costs of electric locomotive electricity can be determined only by analytical methods. Based on the results obtained, the following conclusions were made:

1) A method has been developed for calculating the parameters of the ATM based on experiments of idling and short circuit using computer modeling. Based on the results obtained from the experiment, the following conclusions were made. The analytical parameters in the Mathcad program were determined, which made it possible to improve the performance characteristics and choose the optimal operating modes of asynchronous traction motors.
2) The complex simulation computer model of an AC electric locomotive of a series "O'zbekiston Yo'lovchi" with 4q-S converters and automatic control system has been developed, allowing to conduct research of electromagnetic processes in power circuits.
3) The design parameters and electrical values are determined for the optimal operating modes of the ATM series "1TB2624–0GA02"electric locomotive of the series "O'zbekiston Yo'lovchi".

REFERENCES

[1] N.B. Pirmatov, U.T. Berdiev, K.K. Usmonov, and U.N. Berdierov, "Investigation of the mechanical characteristics of the traction asynchronous motor of the electric locomotive 'O'zbekiston yo'lovchi,'" in the first international scientific and technical conference "railway rolling stock: problems, solutions, prospects" (in Russian) - Tashkent State Transport University (TSTU), 2022, pp. 96-98.

[2] F. Zikrillaev, E.B. Saitov, J.B. Toshov, B.K. Ilyasov, and M.B. Zubaydullayev, "A Software Package for Determining the Optimal Composition and Parameters of a Combined Autonomous Power Supply System Based on Renewable Energy Sources," in AIP Conference Proceedings, 2022, vol. 2432, p. 020021, [Online]. Available: https://doi.org/10.1063/5.0090460.

[3] Y.U.M. Kurbonov, E.B. Saitov, and B.M. Botirov, "Analysis of the influence of temperature on the operating mode of a photovoltaic solar station," ICECAE 2020, IOP Conf. Series: Earth and Environmental Science 614 (2020)012034, [Online]. Available: https://doi.org/10.1088/1755-1315/614/1/012034.

[4] E.B. Saitov, "Optimal model for additional operation of the storage system for photovoltaic wind power plants," E3S Web of Conferences, vol. 220, Sustainable Energy Systems: Innovative Perspectives, SES 2020; Saint-Petersburg; Russian Federation; 29 October 2020, [Online]. Available: https://doi.org/10.1051/e3sconf/202022001080.

[5] E.B. Saitov, Y.B. Sobirov, I.A. Yuldoshev, I.R. Jurayev, and S. Kodirov, "Study of Solar Radiation and Wind Characteristics in Various Regions of Uzbekistan," E3S Web of Conferences, vol. 220, Sustainable Energy Systems: Innovative Perspectives, SES 2020; Saint-Petersburg; Russian Federation, [Online]. Available: https://doi.org/10.1051/e3sconf/202022001061.

[6] E.B. Saitov, "Renewable Energy Development in Uzbekistan: Current Status, Problems and Solutions," RSES 2020, E3S Web of Conferences 216, 01134 2020, [Online]. Available: https://doi.org/10.1051/e3sconf/202021601134.

[7] I. Sapaev, E. Saitov, N. Zoxidov, and B. Kamanov, "Matlab-model of a solar photovoltaic station integrated with a local electrical network," IOP Conference Series: Materials Science and Engineering, vol. 883, no. 1, 012116, IOP Publishing, 2020/7/1. [Online]. Available: https://doi.org/10.1088/1757-899X/883/1/012116.

[8] E.B. Saitov and T.B. Sodiqov, "Modeling an autonomous photovoltaic system in the MATLAB Simulink software environment," AIP Conference Proceedings 2432, 020022, 2022, [Online]. Available: https://doi.org/10.1063/5.0089914.

[9] E.B. Saitov, Sh. Kodirov, B.M. Kamanov, N. Imomkulov, and I. Kudenov, "Increasing the efficiency of autonomous solar photovoltaic installations for power supply of agricultural consumers," AIP Conference Proceedings 2432, 040036, 2022, [Online]. Available: https://doi.org/10.1063/5.0090439.

[10] M. Pustovyetov, K. Soltus, and I. Senyavskiy, “Computer simulation of asynchronous motors and transformers.” Examples of interaction with power electronic converters (in Russian). Monograph. Saarbrucken, Deutshland, Lap Lambert Academic Publishing, 2013. ISBN 3659407763, 9783659407765.

[11] B. Yuldoshov, E. Saitov, J. Khaliyarov, and S. Bobomuratov, "Effect of Temperature on Electrical Parameters of Photovoltaic Module," Proceedings of International Conference on Applied Innovation in IT, vol. 11 (1), pp. 291-295. [Online]. Available: https://doi.org/10.25673/101957.

[12] E. Saitov, S. Akhrorova, O. Juraev, J. Ismailov, and B. Baymirzaev, "Conversion and use of Solar Energy Calculation Methodology for Photovoltaic Systems," Proceedings of International Conference on Applied Innovation in IT, vol. 11 (1), [Online]. Available: http://dx.doi.org/10.25673/101942.

[13] N. Zikrillaev, E. Saitov, J. Toshov, U. Akhmedov, and D. Mukhtorov, "Autonomous Solar Power Plant for Individual use Simulation in LTspice Software Package Booster Voltage Converter," Proceedings of International Conference on Applied Innovation in IT, vol. 11 (1), pp. 207-211. [Online]. Available: https://doi.org/10.25673/101939.

[14] Z. Guo, Q. Wei, and Z. Zhang, "The study on mathematical model and simulation of asynchronous motor considering iron loss," Journal of Physics Conference Series [online], vol. 1060, no. 1, pp. 1-6, 2018. ISSN 1742-6588. Available: https://doi.org/10.1088/1742-6596/1060/1/012085.

[15] A.A. Teryohin and D.A. Dadyenkov, "An overview of how to identify the parameters of an asynchronous electric drive (in Russian)," Vesnik of the Perm National Research Polytechnic University. Electrical Engineering, Information Technology, Control Systems, 2017, vol. 2, pp. 55-66. ISSN 2224-9397/eISSN 2305-2767.

[16] I.I. Epstein, "Calculation of parameters of the asynchronous motor. Energy saving (in Russian)," Energetics. Energy audit, 2014, vol. 8(126), pp. 57-61. ISSN 2218-1849/issn 2313-8890.

Hydraulic and Thermal Engineering Calculation in the Laminar Mode of Operation of a Photoelectric Thermal Battery

Isroil Yuldoshev[1], Shahzod Rahmatillaev[1], Sanjar Shoguchkarov[2] and Uygun Xolov[3]
[1]*Tashkent State Technical University named after Islam Karimov, Universitet Str. 2, Tashkent, Uzbekistan*
[2]*Tashkent University of Applied Sciences, Gavkhar Str. 1, Tashkent, Uzbekistan*
[3]*Karshi engineering-economics institute, Mustaqillik Str. 225, Karshi, Uzbekistan*
yuldashev.i2004@gmail.com, shahzodrahmatillayev12@gmail.com, sanjar42422@gmail.com, uygunshams@mail.ru

Keywords: Hydrodynamic Calculation, COMSOL Multiphysics 5.6, Photovoltaic Thermal Battery, Cooling Methods, Model, Thermal Analysis, Water Velocity, Water Flow, Water Temperature, Pressure, Environment.

Abstract: Efficient conversion of solar energy into electrical and thermal energy has become a major goal of researchers around the world. In this regard, the authors have developed photovoltaic thermal installations to efficiently convert solar energy into electricity and heat. This article briefly analyses the development of a photovoltaic thermal system for efficient cooling of the photovoltaic part with various methods and coolants. A photovoltaic thermal battery (PTB) with a cooling system based on multichannel polycarbonate has been developed. The dimensions of the cellular polycarbonate channels are 7 x 12 mm^2. Water flows horizontally through more than 200 channels in parallel streams. The thickness of the cellular polycarbonate sheet is 4 mm. The PTB cooling system is a structure consisting of a sheet of cellular polycarbonate and channel openings, which are attached to two perpendicularly located polypropylene tubes using transparent silicone sealant. This design of the cooling system (absorber) has less weight and a lower cost compared to traditional metal structures, and the cellular polycarbonate sheet in the PTB is protected from direct exposure to ultraviolet radiation emitted by the sun. The model of a combined PTB installation based on a "photovoltaic battery and heat converter" (PVB-TC) was implemented using COMSOL Multiphysics 5.6. Hydraulic and thermal calculations were carried out in laminar mode, and PTB parameters were determined: water temperature at the outlet of the absorber PTB t_2, water pressure at the inlet of the absorber t_1, and water flow G at the corresponding water velocities W=0.1 m/s, 0.2 m/s, and 0.3 m/s, taking into account three values of ambient temperature - 25 °C, 35 °C, and 45 °C. The modelling process took into account the use of concentrated solar radiation in a combined PEP-TEP installation using weakly concentrating reflectors.

1 INTRODUCTION

The sun, by human standards, is an inexhaustible source of energy. There are devices that convert solar energy into electrical energy. Direct conversion of solar radiation into electrical energy is carried out by photoelectric modules (PEM). An increase in the PEM temperature leads to a decrease in its efficiency.

Therefore, it is necessary to cool the PEM. PEM cooling methods are divided into passive and active, where the coefficient of performance (COP) of polycrystalline silicon PEM depends inversely on the temperature of solar cells (SC) [1-3]. The paper [4] provides an overview of various methods of PEM cooling. However, in most of these cooling methods, the heat generated in the PEM is removed to the environment. Water and air are mainly used as a heat carrier [5-11]. The method of active cooling with water is used in a photo-thermal converter (PTC), which provides consumers with electricity and hot water (thermal energy). Solar cells are usually attached to an absorber plate to improve thermal contact [12–16]. However, Avezov et al. [17] concluded that water-based PTC systems are more efficient than air PTC systems.

Egyptian researchers analyzed all methods of cooling solar modules. Their review includes passive and active cooling methods, cooling with phase change materials (PCM), and cooling of PCM and other additives such as nanoparticles or porous metal [18-21].

The works [22-28] present experimental research of PTC for generating electricity and hot water. The paper [29] considers the electrical and thermal power

of solar hot water systems with single plate collectors to study the mechanisms for determining the output parameters.

The use of polymer and plastic materials in photovoltaic thermal systems has been studied in [30, 31]. Experimental and analytical studies were carried out to study the thermal and energy parameters of PV-T using plastic and polymer materials in plant designs. In [32], the authors conducted a comparative study by modelling PV-T systems of different technologies using the Matlab simulation and ANSYS Software software packages. A reliability study has been conducted on PV-T based on thin-film solar cells based on different technologies (binary, ternary, and quaternary materials). The efficiency obtained by cadmium telluride (CdTe), copper indium diselenide (CIS), and copper indium gallium diselenide (CIGS) PV-T collectors has been found to be more important than the efficiency obtained by silicon and amorphous silicon-based PVT collectors (a-Si), ranging from 47% to 57%. In addition, with this type of PV-T, the outlet temperature of the coolant temperature reaches a value of 43.2°C, which is higher than the value obtained by PV-T collectors based on silicon and amorphous silicon.

In [33], the authors developed an energy model and computer simulation of PV/T for application in buildings. The results show that the overall electrical and thermal efficiencies are 9.39% and 37.5%, respectively.

In this paper, we propose a mathematical model of an installation consisting of a photovoltaic module and an absorber (heat converter) made of a polymer material. The mathematical model of the PEC-THC installation was implemented using the COMSOL Multiphysics 5.6 program.

When cooling the PEC from the back side with water, the surface temperature of the PEC, the amount and final temperature of the cooling water, which, at a sufficient temperature, could be used in the hot water supply system, are of interest.

For the research, a model of a photoelectric thermal battery was chosen, shown in Figure 1, in which a polycarbonate film with channels of a square section 8x8 мм^2 in size was chosen as a cooling element. Cooling water flows through the channels.

2 MATERIALS AND METHODS

For the calculation in COMSOL Multiphysics 5.6, the geometric dimensions of the model and the structure of the PEC materials were specified.

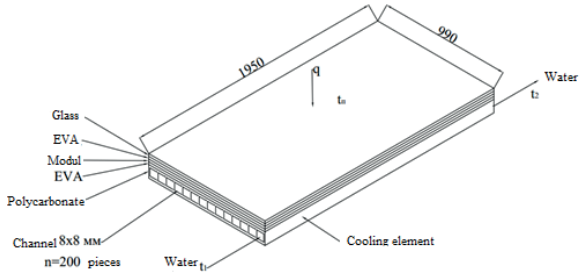


Figure 1: Model for the research of photovoltaic thermal battery.

In addition, the following were asked:

1) Flux density of supplying solar radiation: q, W/m^2.
2) Mode of water flow: Re < 2300.
3) Water inlet temperature: t_1, ºC.
4) Ambient air speed: V, m/s.
5) Barometric pressure: P, bar.
6) Three speeds of water in the channel: W, m/s (0,1,0,2 and 0,3 m/s).

As a result of the calculation under the given conditions, we obtained:

1) Average surface temperature of PEC.
2) Outlet water temperature: t_2, ºC.
3) Inlet pressure of water: P_{vx}, Pa.
4) Water flow through the channel: G, g/s.

3 RESULTS AND DISCUSSIONS

The results of the COMSOL Multiphysics 5.6 calculation are shown in Table 1. The average temperature of the PEC surface is ~45°C (Figure 2). Table 1 shows the results of calculations of the outlet water temperature t_2, the inlet water pressure and the water flow rate G at water velocities W=0.1, 0.2 and 0.3 m/s and at three values of the ambient temperature - 25^0 C, 35^0 C and 45º C.

At water speeds of 0.1, 0.2 and 0.3 m/s, the inlet pressure, or rather the pressure drop, remains constant - 1.8 Pa, and the water flow rate was 1.03, 1.06 and 1.09 g/s, with respectively

Table 1 was used to check for the initial data 1 and Figure 2.

A calculation was carried out to determine the surface temperature of PEC (t_p).

1) The equivalent channel diameter is calculated, as follows:

$$d_e = \frac{4F}{P} = \frac{4 \cdot 8 \cdot 8}{4 \cdot 8} = 8 \text{ mm.}$$

Table 1: Photovoltaic thermal battery simulation results using COMSOL Multiphysics 5.6 at different densities of concentrated solar radiation flux.

Flux density supplying solar radiation [W/m^2]	Inlet water speed t_{ww} [m/s]	Ambient temperature t_a, °C	Inlet water temperature of Absorber t_1 °C	Outlet water temperature of Absorber t_2 °C	Inco-ming water pressure, Pa	Water flow at the absorber outlet, g/s
800	0,1	25	20	24,6	18,8	1,03
		35	20	34,9		1,03
		45	20	43,0		1,03
1000	0,1	25	20	24,7	18,8	1,03
		35	20	34,9		1,03
		45	20	44,6		1,03
1200	0,11	25	20	39,8	18,8	1,03
		35	20	48,2		1,03
		45	20	60,0		1,03
800	0,2	25	20	22,8	18,8	1,06
		35	20	32,9		1,06
		45	20	43,7		1,06
1000	0,2	25	20	22,9	18,8	1,06
		35	20	33,1		1,06
		45	20	43,8		1,06
1200	0,2	25	20	37,7	18,8	1,06
		35	20	46,4		1,06
		45	20	58,2		1,06
		35	20	31,8		1,09
		45	20	40,9		1,09
1000	0,3	25	20	22,7	18,8	1,09
		35	20	32,9		1,09
		45	20	43,4		1,09
1200	0,3	25	20	36,9	18,8	1,09
		35	20	46,1		1,09
		45	20	57,9		1,09
1000	0,1	25	20	24,6	18,8	1,03
		35	20	34,9		1,03
		45	20	43,0		1,03

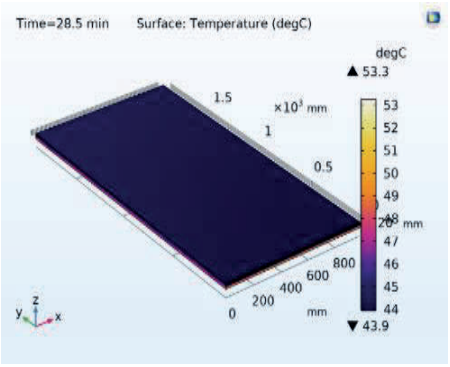


Figure 2: Temperature gradient on the surface of a photoelectric thermal battery made in COMSOL Multiphysics 5.6.

2) Average water temperature determine using:

$$\overrightarrow{t_w}=0{,}5(t_2+t_1)=0{,}5(48{,}2+20)=34{,}1\ ^{o}\text{C}.$$

Surface temperature: t_s=45,9 °C.
Inlet water temperature: t_1=20 °C.
Outlet water temperature: t_2=48,2 °C.
Water speed: W=0,1 m/s.
Water flow: G=1,03 g/s=0,00103 kg/s.
Ambient temperature: t_a=35 °C, at this temperature according to L.2.
Kinematic viscosity: $\mathcal{V}_v$=0,705·10^{-6} m/s.
Density:ρ_v=993,2 kg/m^3.
Dynamic viscosity: μ_v=700,2·10^{-6} Pa·s.

3) We accept wall temperature t_w=40 °C.
At this temperature: μ_w=653,3·10-6 Pa·s.

4) Determining temperature using:

$$t_h=0{,}5(t_{wall}+\overrightarrow{t_w})=0{,}5(40+34{,}1)=37^{o}\text{C}.$$

At this temperature kinematic viscosity:

$$\mathcal{V}_g=0{,}703\cdot10^{-6}\ \text{m/s}.$$

Volume expansion coefficient: β_g=3,67·10^{-4} 1/K.
Heat capacity of water: C_{pg}=4,174 kJ/kg·°C.
Prandtl number: P_{pg}=4,69.
Thermal conductivity: λ_g=0,622 W/m·°C.

5) Determining Reynolds number using:

$$\text{Re}=\frac{W\cdot d_e}{\mathcal{V}_v}=\frac{0{,}1\cdot8{,}0\cdot10^{-3}}{0{,}705\cdot10^{-6}}=1135<2300.$$

therefore the flow regime is laminar.

6) The product of the number Pe (Pecle) and the ratio of the diameter to the length of the channel is determined by

$$\text{Pe}=\frac{d_e}{l}=\frac{4\cdot G\cdot C_{pg}}{\pi\cdot l\cdot\lambda_g}=\frac{4\cdot0{,}00103\cdot4{,}174\cdot10^{3}}{3{,}14\cdot1{,}95\cdot0{,}622}=4{,}51.$$

The product of the reciprocal of Re and the ratio of the length to the diameter of the channel may be calculated using:

$$\left(\frac{1}{Re}\cdot\frac{l}{d_e}\right)=\frac{1}{1135}\cdot\frac{1950}{8}=0{,}215.$$

Correction for the section of hydrodynamic stabilization my be calculated using:

$$\text{Ɛ}=0{,}6\left(\frac{1}{Re}\cdot\frac{l}{d_e}\right)^{-\frac{1}{7}}\cdot\left(1+2{,}5\ \frac{1}{Re}\cdot\frac{l}{d_e}\right)=1{,}148.$$

Channel section: 8x8 mm^2.

7) Rayleigh coefficient (Ra) at temperature t_r=37 °C and is determined using:

$$\text{Ra}=(\text{G}\cdot\text{P}_z)_g=g\cdot\beta_g\ \frac{(t_c-\overrightarrow{t_v})\cdot d_e^3}{\mathcal{V}_g^2}\cdot P_{z_g}=$$

$$=9{,}81\cdot3{,}67\cdot10^{-4}\ \frac{(40-34{,}1)\cdot8\cdot10^{-3}}{0{,}703\cdot10^{-6}}\cdot4{,}69=1\cdot10^{5}<8\cdot10^{5}.$$

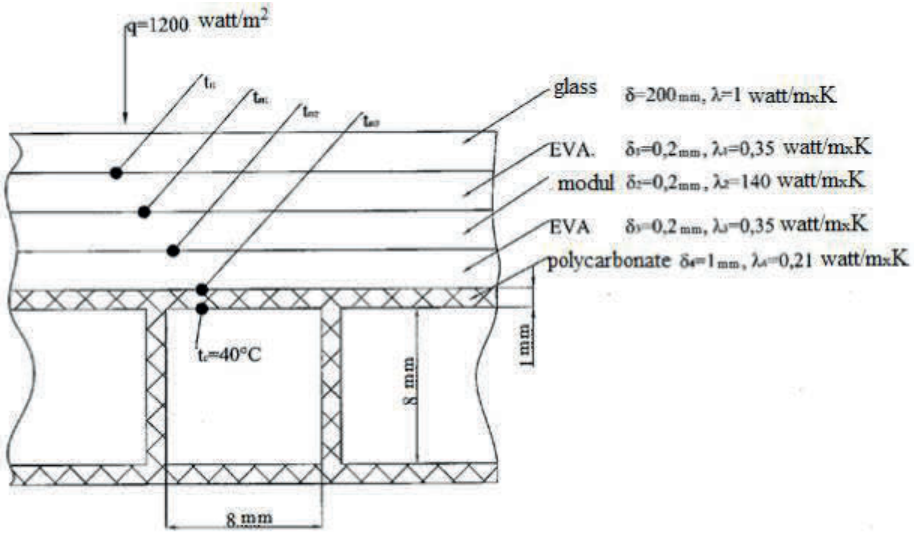


Figure 3: Calculation scheme for determining t_c.

therefore, the flow regime is viscous and according to L.1, it is applicable:

$$\mathrm{Nu_g}=1{,}55\left(Pe_g\frac{d_e}{l}\right)^{\frac{1}{3}}\cdot\left(\frac{\mu_v}{\mu_c}\right)^{0{,}14}\cdot\varepsilon=$$

$$=1{,}55(4{,}51)^{\frac{1}{3}}\cdot\left(\frac{700{,}2}{653{,}3}\right)^{0{,}14}\cdot1{,}148=2{,}97.$$

8) The heat transfer coefficient from the inner wall to the water may be calculated using:

$$\alpha=\frac{Nu_g\lambda_g}{d_e}=\frac{2{,}97\cdot0{,}622}{8\cdot10^{-3}}=230{,}9\ \mathrm{W/m^2\cdot h}.$$

9) The temperature of the inner wall of the channel in contact with water. From Newton's formula may be calculated.

10) $q=\alpha(t_{wall}-\overrightarrow{t_w})$ we find

$$t_{wall}=\frac{q}{\alpha}+\overrightarrow{t_w}=\frac{1200}{230{,}9}+34{,}1=39{,}3\ ^{o}\mathrm{C}.$$

The resulting temperature of 39.3°C is close to the accepted one - 40°C, and then a calculation is made (Figure 3) to determine the surface temperature of the PEC t_c. The surface temperature t_c PEC is calculated without taking into account glass.

The heat transfer coefficient of a photoelectric panel layer is determined using:

$$\mathrm{K}=\frac{1}{\frac{\sigma_1}{\lambda_1}+\frac{\sigma_2}{\lambda_2}+\frac{\sigma_3}{\lambda_3}+\frac{\sigma_4}{\lambda_4}}=\frac{1}{10^{-3}\left(\frac{0{,}2}{0{,}35}+\frac{0{,}2}{140}+\frac{0{,}2}{0{,}35}+\frac{1}{0{,}21}\right)}=169{,}3\ \mathrm{W/m^2\cdot K}.$$

From Newton's formula $q=\kappa\ (t_c-t_{wall})$ we deterine the temperature t_c of the PEC surface:

$$t_c=\frac{q}{\kappa}+t_{wall}=\frac{1200}{169{,}3}+39{,}3=7{,}09+39{,}3=46{,}4\ ^{o}\mathrm{C}.$$

The calculated temperature t_c= 46.4°C is close to the temperature obtained by COMSOL t_c= 45°C.

The agreement between the COMSOL calculation and the heat transfer equations indicates that the COMSOL program is working correctly.

Similar calculations, taking into account the actual water consumption may be calculated using:

$$\mathrm{G=W\cdot}f\cdot\rho=0{,}1\cdot0{,}785\cdot d_e^2\cdot995==0{,}005\ \mathrm{kg/s},$$

show that the surface temperature t_c PEC is 35-36°C and the water temperature at the outlet is ≈30°C.

4 CONCLUSIONS

A mathematical model of a photoelectric thermal battery, consisting of a photoelectric module and an absorber (heat converter) made of a polymer material, is proposed. The mathematical model of the PEC-TEC installation was implemented using the COMSOL Multiphysics 5.6 program and can be used to solve problems of hydrodynamics and heat transfer.

ACKNOWLEDGMENTS

We express our gratitude to the head of the department "AES" of the TSTU named after Islam Karimov, Doctor of Technical Sciences. Yuldoshev I.A. for help in organizing the experiments. The work

was financially supported by the Ministry of Innovative Development of the Republic of Uzbekistan within the framework of the project F-OT-2021-497 "Development of the scientific basis for the creation of solar cogeneration plants based on photoelectric thermal batteries"

REFERENCES

[1] R. Lamba and S.C. Kaushik, "Modeling and performance analysis of a concentrated photovoltaic-thermoelectric hybrid power generation system," Energy Conversion and Management, vol. 115, pp. 288-298, 2016, [Online]. Available: https://doi.org/10.1016/j.enconman.2016.02.061.

[2] I. Jurayev, I. Yuldoshev, and Z. Jurayeva, "Effects of Temperature on the Efficiency of Photovoltaic Modules," Proceedings of International Conference on Applied Innovation in IT, Volume 11, Issue 1, pp. 199-206, doi:10.25673/101938.

[3] V.M. Evdokimov and V.A. Mayorov, "A study of limiting energy and temperature characteristics of photovoltaic solar radiation converters," Applied Solar Energy, vol. 53, no. 1, pp. 1-9, 2017, doi:10.3103/S0003701X17010042.

[4] A. Royne, C.J. Dey, and D.R. Mills, "Cooling of photovoltaic cells under concentrated illumination: a critical review," Solar Energy Materials and Solar Cells, vol. 86, pp. 451-483, 2005, [Online]. Available: http://dx.doi.org/10.1016/j.solmat.2004.09.003.

[5] A. Fudholi, K. Sopian, M.H. Yazdi, M.H. Ruslan, A. Ibrahim, and H.A. Kazem, "Performance analysis of photovoltaic thermal (PVT) water collectors," Energy Conversion and Management, no:78, pp. 641-651, 2014, [Online]. Available: http://dx.doi.org/10.1016/j.enconman.2013.11.017.

[6] A. Ibrahim, A. Fudholi, K. Sopian, M.Y. Othman, and M.H. Ruslan, "Efficiencies and improvement potential of building integrated photovoltaic thermal (BIPVT) system," Energy Conversion and Management, no.77, pp. 527-534, 2014, [Online]. Available: http://dx.doi.org/10.1016/j.enconman.2013.10.033.

[7] A. Ibrahim, M.Y. Othman, M.H. Ruslan, S. Mat, and K. Sopian, "Recent advances in flat plate photovoltaic/thermal (PV/T) solar collectors," Renewable and Sustainable Energy Reviews, vol. 15(1), pp. 352-365, 2011, [Online]. Available: http://dx.doi.org/10.1016/j.rser.2010.09.024.

[8] S.A. Hamid, M.Y. Othman, K. Sopian, and S.H. Zaidi, "An overview of photovoltaic thermal combination (PV/T combi) technology," Renewable and Sustainable Energy Reviews, no.38, pp. 212-222, 2014, [Online]. Available: http://dx.doi.org/10.1016/j.rser.2014.05.083.

[9] R. Kumar and M.A. Rosen, "A critical review of photovoltaic-thermal solar collectors for air heating," Applied Energy, vol.88(11), pp. 603-614, 2011, [Online]. Available: http://dx.doi.org/10.1016/j.apenergy.2011.04.044.

[10] T.T. Chow, W. He, A.L.S. Chan, K.F. Fong, Z. Lin, and J. Ji, "Computer modeling and experimental validation of a building-integrated photovoltaic and water heating system," Applied Thermal Engineering, vol.28(11-12), pp. 1356-1364, 2008, [Online]. Available: http://dx.doi.org/10.1016/j.applthermaleng.2007.10.007.

[11] A. Hazi, G. Hazi, R. Grigore, and S. Vernica, "Opportunity to use PVT systems for water heating in industry," Applied Thermal Engineering, vol.63(1), pp. 151-157, 2014, [Online]. Available: http://dx.doi.org/10.1016/j.applthermaleng.2013.11.010.

[12] T.T. Chow, "A review on photovoltaic/thermal hybrid solar technology," Applied Energy, vol. 87, pp. 365-379, 2009, [Online]. Available: http://dx.doi.org/10.1016/j.apenergy.2009.06.037.

[13] M.A. Hasan and K. Sumathy, "Photovoltaic thermal module concepts and their performance analysis: a review," Renewable and Sustainable Energy Reviews, vol. 14, pp. 1845-1859, 2010, [Online]. Available: http://dx.doi.org/10.1016/j.rser.2010.03.011.

[14] X. Zhang, X. Zhao, S. Smith, J. Xu, and X. Yu, "Review of R&D progress and practical application of the solar photovoltaic/thermal (PV/T) technologies," Renewable and Sustainable Energy Reviews, vol.16, pp. 599-617, 2012, [Online]. Available: http://dx.doi.org/10.1016/j.rser.2011.08.026.

[15] M.Y. Othman, A. Ibrahim, G.L. Jin, M.H. Ruslan, and K. Sopian, "Photovoltaic-thermal (PV/T) technology-the future energy technology," Renewable Energy, vol.49, pp. 171-174, http://dx.doi.org/10.1016/j.renene.2012.01.038.

[16] A. Fudholi, M. Zohri, G.L. Jin, A. Ibrahim, C.H. Yen, M.Y. Othman, and et al., "Energy and exergy analyses of photovoltaic thermal collector with ∇-groove," Solar Energy, vol.159, pp. 742-750, 2018, [Online]. Available: http://dx.doi.org/10.1016/j.solener.2017.11.056.

[17] R.R. Avezov, J.S. Akhatov, and N.R. Avezova, "A review on photovoltaic-thermal (PV-T) air and water collectors," Applied Solar Energy, vol. 47, pp. 169-183, 2011, [Online]. Available: https://doi.org/10.3103/S0003701X11030042.

[18] J.K. Tonui and Y. Tripanagnostopoulos, "Air-cooled PV/T solar collectors with low-cost performance improvements," Solar Energy, vol.81, pp. 498-511, 2007, doi:10.1016/j.solener.2006.08.002.

[19] A. Shahsavar and M. Ameri, "Experimental investigation and modeling of a direct-coupled PV/T air collector," Solar Energy, vol. 84, no. 11, pp. 1938-1958, 2010, doi:10.1016/j.solener.2010.07.010.

[20] F. Sarhaddi, S. Farahat, H. Ajam, and A. Behzadmehr, "Exergetic performance assessment of a solar photovoltaic thermal (PV/T) air collector," Energy and Buildings, vol. 42, pp. 2184-2199, 2010, [Online]. Available: https://doi.org/10.1016/j.enbuild.2010.07.011.

[21] A.S. Joshi and A. Tiwari, "Energy and exergy efficiencies of a hybrid photovoltaic-thermal (PV/T) air collector," Renewable Energy, vol.32, pp. 2223-2241, 2007, [Online]. Available: https://doi.org/10.1016/j.renene.2006.11.013.

[22] A. Bulusu and D.G. Walker, "Review of electronic transport models for thermoelectric materials," Superlattices and Microstructures, vol. 44, pp. 1-36, 2008, [Online]. Available: https://doi.org/10.1016/j.spmi.2008.02.008.

[23] M. Mirzabaev, S.L. Lutpullaev, R.R. Avezov, M.N. Tursunov, A.M. Mirzabaev, and et al., "Patent for utility model “Photoheat converter” No. FAP 00496," Official Bulletin of the State Patent Office of the Republic of Uzbekistan, No. 10(102), pp. 54-55, 2009, [Online]. Available: https://nsp.gov.uz/static/uploads/blutten72c1cd7d-ff60-4dec-84b2-b1b483416e26.pdf.

[24] A.G. Kamilov, R.A. Muminov, and M.N. Tursunov, "Evaluation of solar element and collector system efficiency under hot climate conditions," Applied Solar Energy, vol.44, no. 2, pp. 90-92, 2008, https://www.scopus.com/record/display.uri?eid=2-s2.0-59949089631&origin=resultslist.

[25] R. Santbergen, C.M. Rindt, H.A. Zondag, and R.Ch. Zolingen, "Detailed analysis of the energy yield of systems with covered sheet-and-tube PV-T collectors," Solar Energy, vol. 84, pp. 867-878, 2010, doi:10.1016/j.solener.2010.02.014.

[26] G. Dosymbetova, S. Mekhilef, A. Saymbetov, M. Nurgaliyev, A. Kapparova, and et al, "Modeling and Simulation of Silicon Solar Cells under Low Concentration Conditions," Energies, vol.15,9404,2022, [Online]. Available: https://doi.org/10.3390/en15249404.

[27] S.K. Shoguchkarov, A.S. Halimov, I.A. Yuldoshev, and T.R. Jamolov, "Verification of a Mathematical Model for a Photovoltaic Thermal-Thermoelectric Generator Unit Using Concentrated Solar Radiation," Applied Solar Energy, vol.57, no. 5, pp. 384-390, 2021, doi: 10.3103/S0003701X21050121.

[28] S.A. Kalogirou, "Use of TRNSYS for modeling and simulation of a hybrid PV-thermal solar system for Cyprus," Renewable Energy, vol. 23, no. 2, pp. 247-260, 2001, [Online]. Available: https://www.researchgate.net/profile/Nikhil_Narale/post/I-want-to-model-a-CCHP-system-with-PV-Thermal-cell-as-its-prime-mover-in-trnsys-I-am-new-in-this-program-Can-anybody-help/attachment/5d627830cfe4a7968dc4167c/AS%3A795755832938496%401566734383950/download/2.pdf.

[29] S.A. Kalogirou and Y. Tripanagnostopoulos, "Hybrid PV/T solar systems for domestic hot water and electricity production," Energy Conversion and Management, vol.47, no. 18-19, pp. 3368-3382, 2006, [Online]. Available:https://doi.org/10.1016/j.enconman.2006.01.012.

[30] B. Sandnes and J. Rekstad, "A photovoltaic/thermal (PV/T) collector with a polymer absorber plate: experimental study and analytic model," Solar Energy, vol. 72, no. 1, pp. 63-73, 2002, [Online]. Available: http://dx.doi.org/10.1016/S0038-092X(01)00091-3.

[31] B. Hallmark, C.H. Hornung, D. Broady, C. Price-Kuehne, and M.R. Mackley, "The application of plastic microcapillary films for fast transient micro-heat exchange," International Journal of Heat and Mass Transfer, vol. 51, pp. 5344-5358, 2008, [Online]. Available: http://dx.doi.org/10.1016/j.ijheatmasstransfer.2008.01.036.

[32] H. Haloui, K. Touafek, M. Zaabat, A. Khelifa, "Comparative Study of the Hybrid Solar Thermal Photovoltaic Collectors Based on Thin Films Solar Cells in South of Algeria," Research and Review in Electrochemistry, vol. 8(2), October, 2017, ISSN: 0974-7540.

[33] V. Tomar, G.N. Tiwari, T.S. Bhatti, and B. Norton, “Thermal modeling and experimental evaluation of five different photovoltaic modules integrated on prototype test cells with and without water flow,” Energy Conversion and Management, vol.165, pp. 219-235, 2018, [Online]. Available: https://doi.org/10.1016/j.enconman.2018.03.039.

Protection of Photoelectric Systems from Lightning

Khabibullo Sabirov[1], Tohir Axtamov[1], Sirojiddin Toshpulatov[2], Boysori Yuldoshov[2], Gayrat Raimov[2] and Tulkin Buzrukov[3]
[1]*Physical-Technical Institute, Chingiz Aytmatov Str. 2B, Tashkent, Uzbekistan*
[2]*Termez State University, Barkamol Avlod Str. 43, Termez, Uzbekistan*
[3]*Termiz University of Economics and Service, Farovon Str. 4B, Termez, Uzbekistan*
sabirovhabibullo47@gmail.com, tohiraxtamov@gmail.com, sirojiddin6870@gmail.com, b.yuldoshov10@mail.ru, raimov.gayrat@mail.ru, tolqinbuzrukov@gmail.com

Keywords: Photovoltaic, Photovoltaic Systems, Frequency Inverter, Lightning Protection, Lightning Protection System, Lightning Electro-Magnetic Pulse, Low Voltage.

Abstract: The global PV market has grown extensively for small- to large-scale systems. Inevitably, this leads to the increased development of PV technology. PV systems are intrinsically exposed to weather phenomena. One such phenomenon is lightning. Direct and indirect effects of lightning can cause damage to PV systems. However, lightning protection for PV systems is often neglected and existing standards for protection are underdeveloped. In this paper, previous work is analysed to understand the interaction between lightning and PV systems and to ascertain gaps in current knowledge thereof. Investigation of these aspects will ultimately assist in understanding lightning risk, protection system design and aid in the development of lightning protection standards for PV in the renewable energy industry. The article also presents a specific case study, outlining the lightning protection system of a 3200W photovoltaic plant on a farm. The work includes a thorough description of the lightning protection system, addresses issues related to the contamination of PV surfaces, and proposes practical measures to ensure the safe and reliable operation of photovoltaic systems under adverse weather conditions. The experimental results affirm the significance of implementing lightning protection systems to ensure the robust performance of solar power plants.

1 INTRODUCTION

Nowadays photovoltaics is becoming more widespread due to the reduction in installation costs of photovoltaic (PV) systems and the incentives offered in many countries. Not only private households installing PV systems on their roofs, but private companies, educational institutions, business organizations or farms are also increasingly investing in PV systems to develop unused land. The judicious use of photoelectric batteries (PEB) will give the intended effect when the system is sufficiently focused on dustiness, extremely hot temperatures, and external protection. Factors such as hailstorms, strong winds, and lightning from changing natural phenomena seriously affect the stable operation of the photoelectric stations (PES). Because PEBs cover a large surface area, they are at increased risk of being struck by lightning during thunderstorms [1, 2, 3].

A lightning or lightning protection system consists of external and internal lightning protection measures. It protects people from injury, damage to structures, and electrical equipment from overvoltage.

The installation of photoelectric (PE) modules does not itself increase the risk of lightning strikes to buildings, but in the event of a lightning strike, there may be increased damage to the building's electrical installations. This is based on the fact that due to the wiring of FE systems inside the building (in existing inverters and cable systems), high voltage lightning currents can result in strong conduction voltages and high radiation luminances of lightning. The purpose of the installer is to protect the building and PE systems from damage caused by lightning. Namely, due to a direct lightning strike, electrical and

electronic systems (inverters, remote monitoring systems, the main line of the photoelectric system) must be protected against the effects of lightning electromagnetic impulses (LEMI). Lightning and surge protection of photoelectric devices (PEDs) is actually an understudied topic. The problem of protecting PEDs has been studied from the point of view of preventing direct lightning strikes by constructing conventional external lightning protection systems [4, 5].

However, the problem is more complex; in fact. If a lightning strike occurs in the vicinity of an FE system, the devices may be damaged and fail due to voltage surges and excessive current flows. Previous studies have shown that overvoltages can reach significant values [6] and therefore it is essential to install surge protectors on both sides of power electronic devices (such as charge controllers and inverters) [7].

2 METHODS AND MATERIALS

This article describes the lightning protection system of a 3200W photoelectric plant (PES) installed on a farm in the Tomdi district, Navoi region and the passport parameters of the solar panel used in FES are listed in Table 1.

Table 1: Geometric size and physical parameters of photoelectric battery.

Parameters	
Geometric dimensions	Size
The surface of PEB, S_{PV}	1,018m^2
The frame width of FEB, d	3,5sm
Physical and technical characteristics	
Maximum power of PEB, P_{MAX}	200W
The efficiency factor of PEB, η	19,5%
Open circuit voltage of PEB, U_{OC}	45,48V
Short circuit current of PEB, I_{SC}	7,89A
PEB fill factor of the volt-ampere characteristic, FF	0,74

Tomdi district is one of the districts of Navoi region. It was established on July 3, 1927. It borders Konimekh and Nurota districts in the south, Kyzylorda in Kazakhstan in the north, South Kazakhstan regions in the east, the Jizzakh region in the southeast, and the Uchkuduk district in the west. The center of the district is the village of Tomdibulok. The territory of the district is located in the Kyzylkum desert, the geographical coordinates are latitude 42° 13' 30" North, longitude 65° 14' 5" East, and the height of the land above sea level is 220 m (Figure 1 and Figure 2).

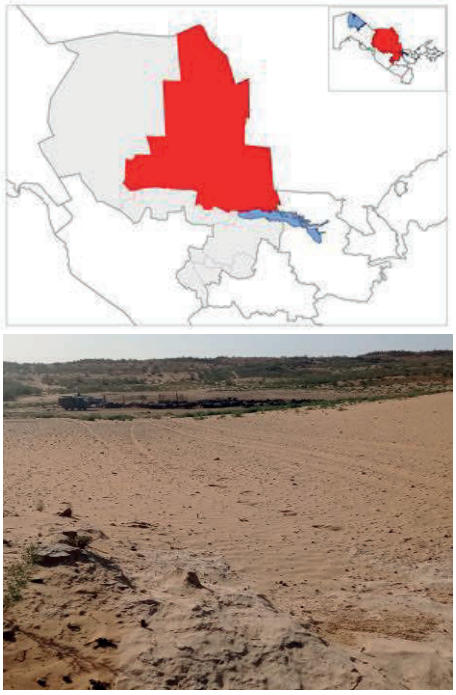


Figure 1: Geographical map of the area.

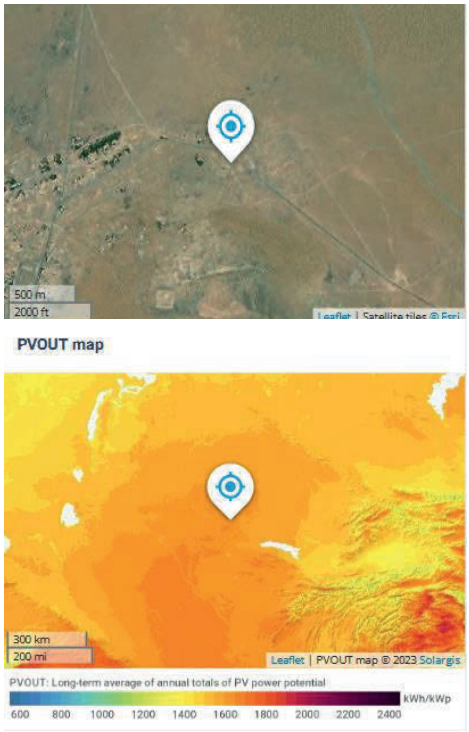


Figure 2: Solar Atlas data of the region.

The total area of Tomdi district is 4248.5 thousand km^2, of which 30.3 thousand km^2 are pastures (71%) and 12.2 thousand km^2 (29%) are non-agricultural lands. There are no irrigated lands. The climate is strictly continental. The average temperature in January is -4.1°, and the lowest temperature is -31°. The average temperature in July is +30°, the highest temperature is +48°. The average annual precipitation is 108 mm. There are no rivers or lakes. The soils are red, sandy soils.

The photoelectric plant installed in this area was fixed to the structure fixed to the ground, and before the installation work, information related to the area was collected and analyzed through the [8].

According to it, the azimuth angle for the solar panels was 180 ° clockwise from the north, and the optimal value of the installation angle of the solar panels with respect to the ground was 36°. It was determined that the total annual energy produced by PES is 5,140 MW•h and the annual total solar radiation is 2,002.2 kW/m^2.

Annual solar radiation values for the area are given below; direct normal radiation - 1772.2kWh/m^2, total horizontal radiation - 1708kWh/m^2, diffuse horizontal radiation - 654.6kWh/m^2, total radiation falling at an optimal angle – 2010.4 kWh/m^2 was found to be. Figure 3 shows the graphs of solar radiation and power that PES can produce per month and hours.

Figure 3: Power and Solar radiation data of the installed station.

3 RESULTS AND DISCUSSION

The farm is located in a desert area 352 km from the center of the district. This cattle breeding farm mainly uses PES to water its animal husbandry twice a day. For watering works, water is filled 3 times a day from a 65 m deep well into a cistern with a capacity of 8m^3 [9]. This year, some changes were made to the installation work of PES, which was installed on the farm last year. Accumulators of autonomous PES cannot justify themselves due to the sharp increase in climate in desert zones for the summer season. Therefore, the battery system was changed to a system that works only during the day with the help of a frequency inverter, which is shown in Figure 4.

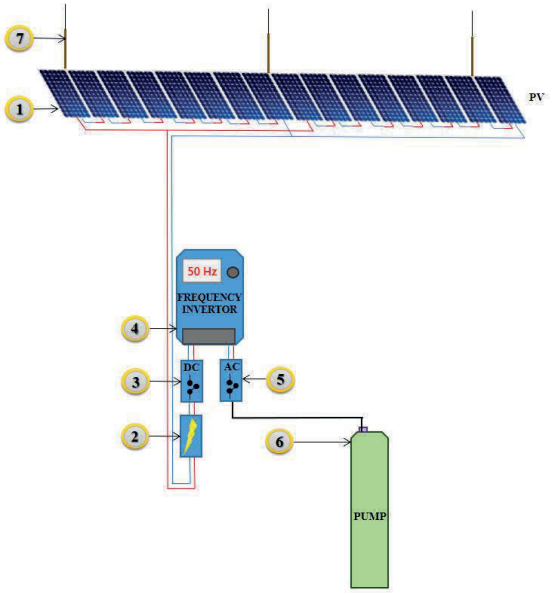


Figure 4: Drawing of a photoelectric plant for pumping water from a well through a frequency inverter installed on a farm: 1) mono silicon photoelectric battery, 2) lightning protection for electronic equipment (KYV-PS,1000DC/32A), 3) constant current connector, 4) frequency inverter (USFULL, 2,2kWt), 5) alternating current connector, 6) well pump (ZEN, 750 Wt), 7) protection of photoelectric station through the ground.

The changes made to the mounting work are as follows: for the external lightning protection system of the FES in the border area, vertical metal poles with an interval of 4 m each were installed on the ground at a distance of 1.5 m from the system. These poles protect photovoltaic cells on the surface, photovoltaic cell profiles, and structural parts from

lightning strikes. Kayal KYV-PS (1000DC/32A) brand and IEC (EN) 60269-6 series European standard lightning protection device was used to protect the electronic components before the constant current from the panels enters the inverter (Figure 5).

Figure 5: Overview of lightning-protected PES and electronic devices.

It is known that during the PEB operation, the surface is polluted due to dust in the air, and the dust settles on the surface in certain microns. Due to the variability of the weather, especially during the rainy season, the sky is covered with clouds and it rains, which washes away the dust on the surface of the PEBs and cleans the panels. Naturally, in this case, the electrical efficiency of photoelectric batteries increases, and in addition, the value of the induced voltage associated with the magnetic fields of lightning flashes increases, and voltage loads are observed in electronic devices. This situation causes several problems in the operation mode of electronic devices and equipment. Our experience shows that it is necessary to complete the stage of lightning protection before the full operation of any PES installed on the premises. In addition, it is important that the panel frames of PEBs are painted in black to protect against lightning [10].

This work of photoelectric module protection systems was performed with the intention of contributing to providing experimental data for improved cases and this work is part of ongoing research.

4 CONCLUSIONS

During the works carried out in the area, the necessary installation works for optimal operation of the autonomous photoelectric plant for raising water from a well with a depth of 65 m were performed. The problems of providing the farm with electricity and water necessary for the whole year and especially for the summer season in livestock farms during the extreme months, which are the most unfavorable in terms of climate, were studied and solved. In addition, special attention was paid to the protection of the photoelectric station from adverse weather conditions. To protect against unexpected overloads, PES and consumer devices were protected according to world standards. The head of the farm acknowledged that the system was justified after a quarter of the work process of the station. Based on the results of the experiment, the following can be concluded:

1) Lightning strikes can seriously damage solar panels, inverters, electrical systems and monitoring equipment.
2) Introduction of lightning protection systems allows solar power plants to deliver clean and reliable energy to the network even in adverse weather conditions and ensures continuous operation.
3) In addition to extensive grounding measures, it is important to consider an isolated location on high ground from a high lightning zone, dry, rocky or otherwise poorly conductive soil layers, and the radius of the grounding connection.

REFERENCES

[1] E. Collins, M. Dvorack, J. Mahn, M. Mundt, and M. Quintana, “Reliability and availability analysis of a fielded photovoltaic system,” in Proc. 34th IEEE

Photovoltaic Specialists Conference (PVSC), 2009, pp. 2316 - 2321.

[2] G. Kálecz, Z. Tóth, N. Szedenik, B. Németh and I. Kiss, "Zone-Based Lightning Protection Method for High-Performance Photovoltaic Power Plants," 2021 35th International Conference on Lightning Protection (ICLP) and XVI International Symposium on Lightning Protection (SIPDA), Colombo, Sri Lanka, 2021, pp. 1-5, doi: 10.1109/ICLPandSIPDA54065.2021.9627397.

[3] A. Rousseau and M. Guthrie, "Risk assessment of lightning-induced surges on PV Systems," In International Colloquium on Lightning and Power Systems, 2017, pp. 1-8.

[4] G. Kálecz, Z. Tóth, I. Kiss, and B. Németh, "Theory behind the Zone Concept for External Lightning Protection of Photovoltaic Power Plants," Electric Power Systems Research, vol. 209, 2022, pp.108025.

[5] C. B. Rogers, "The protection of photovoltaic power systems from lightning," in Proc. 15th IEEE Photovoltaic Specialists Conf., 1981, pp. 761-766.

[6] International Standard, "Overvoltage Protection for PV Power Generating Systems-Guide," IEC Std. 61173.

[7] P. Vangala, M. Ropp, K. Haggerty, K. Lynn, and W. Wilson, (2008). "Field measurements of lightning-induced voltage transients in PV arrays," in Proc. 33rd IEEE Photovoltaic Specialists Conf., 2008, pp.1 - 4.

[8] Global Solar Atlas, [Online]. Available: https://globalsolaratlas.info.

[9] R. A. Muminov, E. Z. Imamov, M. N. Tursunov, Kh. Sabirov, and T. Z. Akhtamov, "Photovoltaic installation for lifting water from deep wells under extreme conditions," Republican Scientific and Practical Conference on Current Issues in the Physics of Semiconductors and Polymers, Tashkent, NUU, 2022, pp. 179-180.

[10] H. J. Stern and H. C. Karner, "Lightning induced EMC phenomenain photovoltaic modules," in Proc. IEEE Int. Symp. Electromagnetic Compatibility, 1993, pp. 442-446.

Steam Gas Plant Reducing Circulating Water Waste in Water Cooling Towers

Qobilbek Dadaboyev[1,2] and Bakhtiyar Yunusov[1]
[1]*Tashkent State Technical University, Universitet Str. 2, Tashkent, Uzbekistan*
[2]*University of Tashkent for Applied Sciences, Gavhar Str. 1, Tashkent, Uzbekistan*
qobilbekturacity@gmail.com, elyor.saitov@mail.ru

Keywords: Experimental, Disaggregation, Cellulose, Concentration.Technical Water, Cooling System, Ventilator, Water Trap, Water Spray Nozzle, Water Vapor, Step Water Trap, Water Treatment.

Abstract: This work is devoted to the calculation of the cooling tower. The explanatory note reflects the material, thermal, hydraulic calculations, guided by which it is possible to select the type of apparatus and its design dimensions. The temperature diagram of the process, the technological diagram of the process, the structural diagram of the apparatus, sketches of the main elements of the apparatus are also shown. The field of velocities of the gas phase, the movement of solid particles, the kinetics of disaggregation of lumps of material, the effect of the concentration of the inert packing and the swirl angle of the coolant on the kinetics of disaggregation of the coolant, was studied experimentally and the choice of the type of inert substance and its concentration was determined. With the help of these techniques how much technical water is saved is also given. In addition, the kinetics of material drying was studied taking into account its disaggregation in a swirling coolant flow.

1 INTRODUCTION

Water cooling in industry Industrial water supply systems are designed to supply production with water in the required quantity and appropriate quality [1]. They consist of water intakes, pumping stations, interconnected structures of water conduits, water quality treatment and improvement facilities, control and storage tanks, water coolers, and distribution pipelines [2]. Water is stored for reuse after cooling and (or) cleaning. Such water is called recycled or circulating. Depending on the type of technological process, it can be a circulating water transport or absorbing medium (the use of water of such qualities is not taken into account in this work) or a circulating heat carrier in the cooling system of the circulating water supply [3]. It is a system in which water is used as a refrigerant to cool equipment or to condense and cool gaseous and liquid products in heat exchangers, which are heated and in some cases contaminated with these products, mainly due to leaks in water pipes. Basically, after cooling and cleaning (if necessary) in cooling towers, the main part of the water is returned to the system; a part of the processed water (usually no more than 5%) is lost to the effluent in the form of evaporation, droplet penetration, leakage and system bursting. In the world, recycled cooling water is used to cool various types of technological equipment, which is about 65% of the total water consumption of this category on average for all industries. and is determined by the operational characteristics of the equipment. When choosing the type of cooling towers to ensure this temperature, it is necessary to take into account the possibility of water contamination with production products in the water cycle. Heat and power industry enterprises consume two-thirds of fresh water from water supply sources for industrial needs, and the largest consumption is for cooling technological equipment (96%). At the same time, the water turnover ratio in the industry is lower than the average for the industry and is approximately is 60%. In industry and energy, water is used to condense and cool gaseous and liquid products [4].

Water cooling tower – serves to lower the temperature of "Technical water" participating in the process of condensation in the cooling system along the general cycle and after returning from the cycle. It removes heat from circulating water with a temperature of 26-28 ^{0}C and transfers this heat to the incoming air flow, as a result, it provides the cooling system with technical water with a constant temperature of 16-18 ^{0}C.If we look at the world experience, it seems that there are different styles in

the technical water cooling system; with outdoor pool, spray pool and cooling tower. Among these methods, the most optimal choice from the economic and environmental point of view is the cooling tower method, because it does not require a large area like the other methods, and the water cooling capacity and quantity are large, and the work efficiency is high. Therefore, in recent years, countries around the world have been using only the cooling tower method [5]. Now let's look at the technical parameters of the air-foil cooling tower from the beginning. The top of the radiator block has circular air intake fins that pull up the warm air below. 2 of these tunnels serve a total of 22 tunnels, 11 of them for a steam gas unit with a capacity of 450 mw [6].

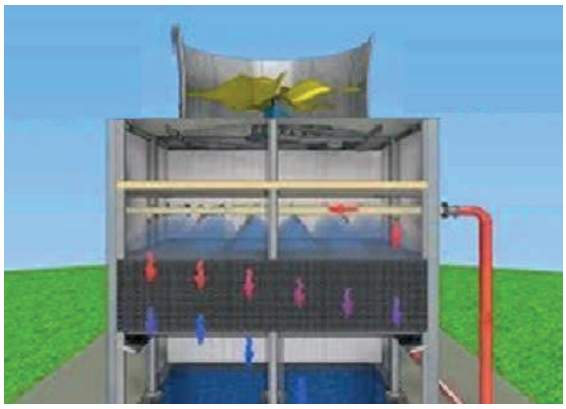


Figure 1: Water cooling tower.

The power of the engine turning the blades is 199 kW/h, Voltage- 400 V the shaft that transmits rotational movement from the engine is 159 mm.

The hot water coming from the common system is transferred through pipes to the sprinklers in the inner, upper (lower than the wings) part of the greenhouse. Sprayers spray hot water on the rust in the form of a drop (spray) [7]. Sprayed hot water droplets hit *the apacume l* and spread downwards. At the bottom of the terrace there is a water pool where the cooled water is stored. Cold water is sent back to the system through supply pumps and circulation continues.

2 METHODS AND MATERIALS

This research work differs from the works in the literature mentioned above. Two identical mono c-Si PV modules were used in the experiment. The electric power of PV modules is 15W and consists of 36 SC. The PV modules are mounted on a support device that has a two-axis twist at the same angle. The base device also has a place for measuring instruments (Figure 2) [8].

Water from the outside pipe is distributed to each distribution pipe and then to the sprinklers.
Distribution pipe material - FRP (Fiberglass).The Figure 2 shows a typical water cycle around a cooling tower. They receive circulating water cooled through these heat exchangers. During this process, heat is absorbed into the circulating water and returned to the cooling tower for cooling. This process is continuous while the station is running. Make-up water is supplied to the cooling tower pool to compensate for evaporation losses, displacement losses, etc.

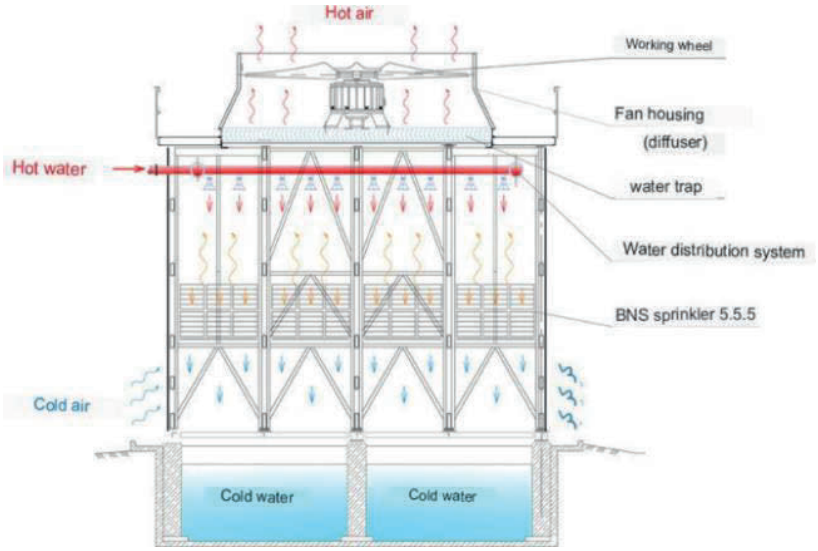


Figure 2: Mode of operation of the water cooling tower.

A bypass line is used in the winter season [9]. A water cooling tower can lose 2% of the water it cools during operation. For example, if a cooling tower that cools 35,000 m^3 of water per hour wastes 700 m^3 hour of water, it becomes 16,800 m^3 per day. Since this is definitely chemically purified, processed water, its price is quite high. Let's take a look at exactly where this waste of water is taking place. The hot water coming from the system to the water cooling tower is sprayed downwards from the nozzles of the tower (Figure 2), and the air fins suck the air up, as a result, the hot water particles sprayed from the nozzles fall on the arasitel and begin to flow downward [10]. The upward movement of air and the downward movement of water along the air duct creates a mutual heat exchange process, and the air takes the heat of the water and goes out into the atmosphere through the ventilator. At first glance, everything is normal, because the goal has been achieved, the temperature of the production technical water has decreased, but during this process, a part of the water that has not yet reached the surface is released into the atmosphere with the hot air due to the force of the air flow. Of course, water catchers are installed under the wings, which will catch some of the water droplets, but to catch more water, we need a different technology. We have concluded from our research that to reduce this wastage, wrapping stepped catchment panels around the tower will reduce wastage to a certain extent.

Step water catchment device, first of all, let's take a good look at the problem before us.

3 RESULTS AND DISCUSSION

Technical water lost in the cooling tower is lost in the form of water vapor and water particles. We now aim to reduce the wastage of water particles rather than water vapor retention. Because trapping water vapor has some complications. The waste of water particles can be reduced by technical methods. To do this, we need to install water-catching steps on the upper part of the ventilator, which do not block the air flow, but catch water particles. This device of ours should be such that it traps water particles and eats them together, but does not prevent the air from escaping. The water trap with steps that we recommend has the same capabilities as the steps. This device is installed tightly on the top of the ventilator [11].

Installation of a step water catchment structure. Our recommended step water trap must meet several requirements.

1) The performance of the device justifies itself.
2) It should have a soda texture as much as possible.
3) Not requiring energy in working condition.
4) Resistance to mechanical deformation.
5) Chemical corrosion resistance.
6) Do not disturb the air outlet so as not to stress the fan motor.

For emergency situations, the on son can be released and separated into pieces. Of course, there is a risk that salt particles will collect and form slag layers between the steps. Therefore, it can be released for cleaning purposes [12].

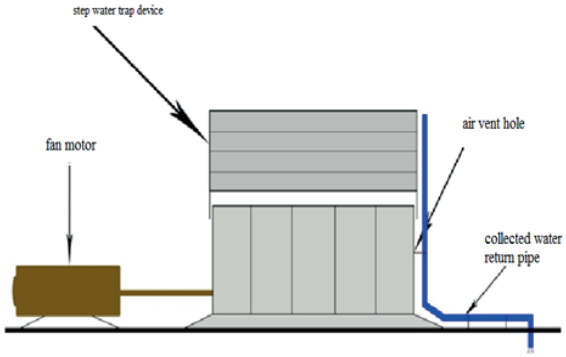


Figure 3: Installation of the device.

Now we will approach each requirement separately.

Step water catchment device is reminiscent of gas turbine blades (Figure 3). That is, it consists of panels with a special water catchment channel consisting of 4 steps.

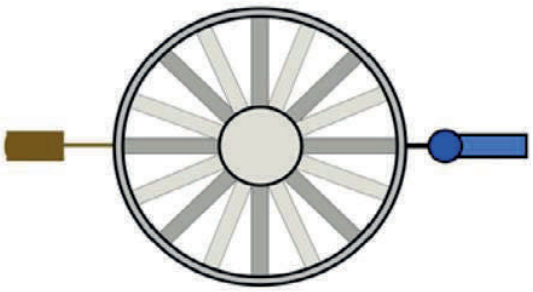


Figure 4: Top view of a stepped water trap.

Is 970 cm, corresponding to the diameter of the ventilation hole. Correspondingly to the dome of the tunnel, the contact block is bricked with 8 fixing bolts of the panel water channel so to structure the owner she is slip coming the water internal channel across skip sends. O is fixed water particles collect and move along the channel. Because the panels are installed at

a 30° slope along the center of the circle. Then the water collected in the water channel begins to flow towards the edge of the circle (Figure 4). The water moving along the water channel goes around the edge of the panels to the bricked water pipe. Through the central water return pipe, water is poured into the cooling tower water supply. The efficiency of water heaters depends on the speed of air ώ flow and the density of water particles q_w. We can express this relation as follows:

$$q_v=3{,}44\cdot10^{-5}q_{tv}\,q_w{}^x\acute{\omega}^y$$

or to simplify, on the basis of legitimacy

$$q_v=^5 q_{tv}\,(q_w\acute{\omega}/14{,}57)^y,$$

here q_{tv} -table values, q_v at ώ= 2,35 m/s va q_w=6,2m^3 /(m^2·h) [13] the performance of water catchers is not universal for any amount of air speed. In addition, the granulomere composition of water particles in the air greatly affects the performance.

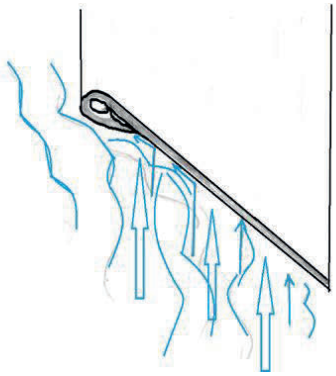


Figure 5: Cross-sectional surface of the panel.

The panels of our device have such a structure that it is installed under a special slope of 70°, and after the water particles hit the surface of the panel, they move up along the panel due to the air flow and flow into the special water channel at the top of the panel and move along the water channel (Figure 5 and Figure 6).

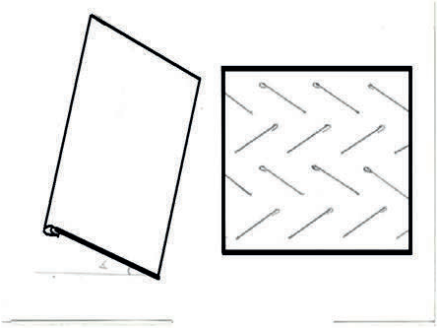


Figure 6: Panels step mothers across location.

And the air flow continues to slide up along the upper part of the channel without entering the channel due to the almond-shaped structure at the entrance of the channel, and after 4 steps, it goes out into the atmosphere (Figure 7).

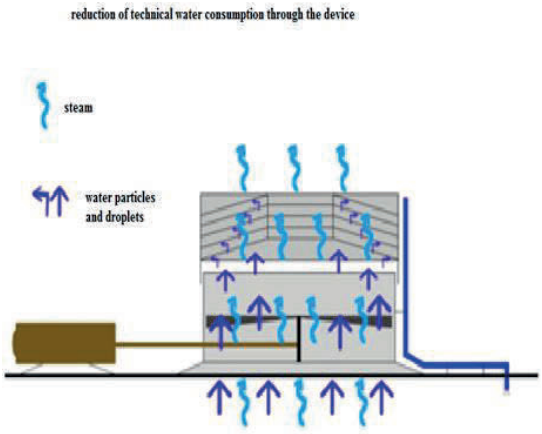


Figure 7: Water retention in the device.

There are 16 panels parallel to each other in the horizontal direction, and 4 steps in a zig-zag pattern in the vertical direction. The total number of panels reaches 64 in one device.

4 CONCLUSIONS

Then it becomes possible to hold more water particles in the air stream [14-16].

Our device now only captures water particles. And the water vapor goes out into the atmosphere with the air flow. Our device must be made of corrosion-resistant material, as it will work with constant water. There are many types of such materials today, but the most economically optimal choice is FRP (glass fiber reinforced concrete) cooling pipes. In addition, FRP is mechanically as strong as steel due to its high inter-lattice uniformity. The price is not expensive, that is, it is cheaper than steel. If we install a total of 22 such devices for a common cooling tower, we will catch 65-70% of the lost technical water particles and droplets. That is, 50% of the total loss of 600 m^3 /h is steam, 50% is in the form of water particles and drops, and 300 m^3 is in the form of water particles and drops. 65 % of technical water retained is 195 m^3. If we put it in money, the amount saved every hour is 8,775,000. The amount lost per hour was 27,000,000. This is 32.5% of the total amount.This means 32.5% of the waste lost through the vent hole. If this device is

introduced in our production, our device will definitely justify its cost and the station will achieve more economic efficiency.

REFERENCES

[1] B. Yunusov and M. Asimova, "Creation of Hybrid Solar - Thermal Installations in the Climatic Conditions of Uzbekistan," AIP Conference Proceedings 2552, 050032, 2022, p. 113491, [Online]. Available: https://doi.org/10.1063/5.0130121.

[2] B. Yunusov and M. Tazhiddinova, "Research of the processes of drying, disaggregation and separation of powder cellulose in one apparatus," AIP Conference Proceedings 2552, p. 050033, 2023, [Online]. Available: https://doi.org/10.1063/5.0130122.

[3] B. Yunusov and J. Normuminov, "Some features of the grinding process in the production of powder materials," E3S Web of Conferences 216, 01131, 2020, p. 160-167, [Online]. Available: https://doi.org/10.1051/e3sconf/202021601131.

[4] F. Zikrillayev, E. B. Saitov, J. B. Toshov, B. K. Ilyasov, and M. B. Zubaydullayev, "A Software Package for Determining the Optimal Composition and Parameters of a Combined Autonomous Power Supply System Based on Renewable Energy Sources," in AIP Conference Proceedings, 2022, vol. 2432, p. 020021, [Online]. Available:

[5] Y. M. Kurbonov, E. B. Saitov and B. M. Botirov, "Analysis of the influence of temperature on the operating mode of a photovoltaic solar station," ICECAE 2020, IOP Conf. Series: Earth and Environmental Science 614, 2020, p. 012034, [Online]. Available: https://doi.org/10.1088/1755-1315/614/1/012034.

[6] E. B. Saitov, "Optimal model for additional operation of the storage system for photovoltaic wind power plants," E3S Web of Conferences, Sustainable Energy Systems: Innovative Perspectives, SES 2020; Saint-Petersburg; Russian Federation; 29 October 2020, vol. 220, [Online]. Available: https://doi.org/10.1051/e3sconf/202022001080.

[7] E. B. Saitov, Y. B. Sobirov, I. A. Yuldoshev, and S. Jurayev, "Study of Solar Radiation and Wind Characteristics in Various Regions of Uzbekistan," E3S Web of Conferences, Sustainable Energy Systems: Innovative Perspectives, SES 2020; Saint-Petersburg; Russian Federation; vol. 220, 29 October 2020, [Online]. Available: https://doi.org/10.1051/e3sconf/202022001061.

[8] E. B. Saitov, "Renewable Energy Development in Uzbekistan: Current Status, Problems and Solutions," RSES 2020, E3S Web of Conferences 216, p. 01134, 2020, [Online]. Available: https://doi.org/10.1051/e3sconf/202021601134.

[9] I. Sapaev, E. Saitov, N. Zoxidov, and B. Kamanov, "Matlab-model of a solar photovoltaic station integrated with a local electrical network," IOP Conference Series: Materials Science and Engineering, 883, 1, p. 012116, IOP Publishing, 2020/7/1, [Online]. Available: https://doi.org/10.1088/1757-899X/883/1/012116.

[10] J. Toshov and E. Saitov, "Portable autonomous solar power plant for individual use," E3S Web of Conferences 139, p. 01087. 2019, [Online]. Available: https://doi.org/10.1088/1757-899X/883/1/012116.

[11] E. B. Saitov and T. B. Sodiqov, "Modeling an autonomous photovoltaic system in the MATLAB Simulink software environment," AIP Conference Proceedings 2432, p. 020022, 2022, [Online]. Available: https://doi.org/10.1063/5.0089914.

[12] E. B. Saitov, Sh. Kodirov, B. M. Kamanov, N. Imomkulov, and I. Kudenov, "Increasing the efficiency of autonomous solar photovoltaic installations for power supply of agricultural consumers," AIP Conference Proceedings 2432, p. 040036, 2022, [Online]. Available: https://doi.org/10.1063/5.0090439.

[13] F. Zikrillayev, E. B. Saitov, J. B. Toshov, B. K. Ilyasov, M. and B. Zubaydullayev, "A software package for determining the optimal composition and parameters of a combined autonomous power supply system based on renewable energy sources," AIP Conference Proceedings 2432, p. 020021, 2022, [Online]. Available: https://doi.org/10.1063/5.0090460.

[14] Y. Su, H. Zhao, X. He, Z. Zheng, Q. Ma, and J. Ding, "The effect of wet-grinding phosphorus slag on the hydration kinetics of Portland cement," Construction and Building Materials, vol. 364, 18 January 2023, p. 129942, [Online]. Available: https://doi.org/10.1016/j.conbuildmat.2022.129942.

[15] M. C. V. Bravo, M. J. S. Lara, M. C. Padilla, M. E. P. Ramos, and S. J. Ambriz, "Effect of wet grinding and drying of the nixtamal on physicochemical, morphologic, and vibrational properties of flours and rheological changes of masa" February 2021, Journal of Food Processing and Preservation 45(5), May 2021, p. e15415, [Online]. Available: https://doi.org/10.1111/jfpp.15415.

[16] A. I. Ol'shanskii and A. N. Golubev, "Investigation of the Kinetics of Heat and Moisture Exchange during Heat Treatment and Drying of Thin Wet Thermal Insulation Materials" February 2023, ENERGETIKA Proceedings of CIS higher education institutions and power engineering associations vol. 66(1), pp. 66-79, [Online]. Available: https://doi.org/10.21122/1029-7448-2023-66-1-66-79.

The Development of Sensitive Measuring Schemes for Capacitive-Semiconductor Humidity Transmitters

Anvar Rakhmanov[1], Stepan Kuznetsov[1,2] and Feruz Fattoev[1]
[1]*Tashkent State Technical University, Universitet Str. 2, Tashkent, Uzbekistan*
[2]*University of Tashkent for Applied Sciences, Gavhar Str. 1, Tashkent, Uzbekistan*
elyor.saitov@mail.ru

Keywords: Measuring Circuit, Humidity, Capacitive-Semiconductor Converter, Accuracy, Bipolar Voltage, Pulse Duration.

Abstract: Capacitive-semiconductor converters play a crucial role in converting physical quantities with complex parameters due to their positive properties, including measurement accuracy, reliability, manufacturability, and efficient measurement schemes. Experimental results and discussions reveal that at small and medium humidity values (2÷6%), the bridge measuring circuit may not provide the necessary gradation and sensitivity of the output signal. To address this, a measuring circuit is developed to increase sensitivity and linearize the output characteristics of the capacitive-semiconductor converter at low and medium humidity values. The proposed measuring circuit utilizes a bipolar voltage for input value modulation, introducing a new construction principle for capacitive-semiconductor humidity converters. The measuring scheme enhances the accuracy and linearity of static characteristics by utilizing bipolar pulses of varying durations. The study establishes the practical value of pulsed power supply, providing additional opportunities for linearization of conversion characteristics. The presented results contribute to the development of effective measuring circuits for capacitive-semiconductor converters, enhancing their metrological characteristics and widening their applications in monitoring and control systems.

1 INTRODUCTION

In information-measuring equipment capacitance - semiconductor converters of humidity, flow, concentration and temperature of various objects and media are widely used to convert physical quantities with complex parameters, which is explained by a number of their positive properties, such as the accuracy of measurement of the parameters of objects, reliability, manufacturability and efficiency of measurement schemes [1].

The basis of the principle of action of capacitive-semiconductor converter of dispersed media humidity is the dependence between the humidity of the measured medium W and its electro-physical properties, i.e. the change in the active and reactive resistance of the transducers depending on the dielectric constant of the wet medium. The essence of the method is to determine the permittivity of the medium by measuring the active and reactive resistance and energy losses of the electric capacitor, in which the role of the dielectric is played by the media [2].

The capacitive humidity converter was created taking into account the fact that at the same time the temperature, the total content of soluble salts and the humidity of the dispersed medium will be determined at the same point. Consider capacitive semiconductor transducer is a transducer in the form of a probe, allowing measurement in dispersive media without extracting the samples [3].

The design of the capacitance-semiconductor converter is performed so that their longitudinal length is almost an order of magnitude greater than the transverse dimensions and therefore the flow distribution calculations should only take into account changes in the field in the longitudinal direction. The static and dynamic characteristics of the probe capacitance-semiconductor converter can be analyzed in detail on the basis of a sensible mathematical apparatus describing the law of change in the flow distribution along the path of the flow lines, flow distribution changes $I_i(x)$, voltages $U_i(x)$ and the laws of change in the inverse reduction functions [4].

Graph models of the distributed circuit of the capacitive-semiconductor converter are reduced to the transition from complex differential and integral equations to the discretization of algebraic relations, describing and displaying unknown values of the quantities that are the nodal element in the distributed circuit of the converter. When studying and displaying the transformation circuit of the probe capacitance-semiconductor converter based on the graph model, this circuit is conditionally divided into elementary sections Δx, consisting of input and output quantities and parameters, which, in turn, are exposed to external influences of quantities of different physical nature. In this case, the laws of change and distribution of external influencing quantities may be different, because the change in the laws of distribution depends on the physical nature of the chain, where the corresponding parameters are produced [5-8].

2 METHODS AND MATERIALS

Statistical characteristics of a capacitance-semiconductor Converter with distributed parameters can be determined directly from the structural schemes and on the basis of a graph model using descriptions of inter-chain and intra-chain effects, as well as rules for finding the output values of serial, parallel and mixed connections of several transducers. In cases where the use of physical effects or intra-chain conversion for the implementation of the respective devices is required to provide the distribution in space of the parameter [9].

In capacitance-semiconductor converters, the input value is associated with the modulation of the electric flow F_{el}, which is a new principle of construction of capacitive-semiconductor humidity converters and is used to modulate humidity in the ranges $w_{\min} \le w \le w_{\max}$. To implement the described principle, a bipolar voltage is used as a reference source. It should be noted that when the input parameters are modulated with a change in the circuit function $\eta_i\left(w_{inp.}\right)$, the transformation parameter changes λ_i. However, you can find a range where the conversion parameter change λ_i is negligible, a schematic function (1) that can be represented as [4]:

$$\eta_э\left(w_{inp.}\right) = \eta_{э0}\left[e^{w\left(w_{inp.}\right)+\varphi_0} + 1\right]. \quad (1)$$

The static characteristic of the capacitance-semiconductor Converter (2) based on the developed generalized graph model for converting the humidity of dispersed media has the form [10]:

$$U_2 = U_1\eta_1\lambda_1 K_1\left(I_1, I_p\right)K_2\left[w_{inp.}\,\eta_p\left(w_{inp.}\right)\lambda_p\left(w_{inp.}\right)\right]K_3\left(U_p, U_2\right)w_{inp.}, \quad (2)$$

where, U_1 – the input voltage; U_2 – output voltage; λ_i - parameter conversion of the input value; K_1, K_2, K_3 – inter-chain coefficients; $\eta_i\left(w_{inp.}\right)$-circuit function.

Depending on the type of capacitance-semiconductor Converter and its geometric parameters, there may be different laws of distribution of specific reactive and active conductivity, which is described by the circuit function and further on the basis of simulation using a graph model, a feature of the distribution of specific reactive conductivity, which depends on the humidity of the dispersed medium, is revealed.

3 RESULTS AND DISCUSSION

It is established that at small and medium values (2÷6%) of the dispersed media humidity, the bridge measuring circuit does not provide the necessary gradation and sensitivity of the output signal. To increase the sensitivity of the output signal and linearization of the output characteristics of the capacitance-semiconductor converter at low and medium humidity values, a measuring circuit is developed (Figure 1a), allowing with high enough accuracy to convert small changes in the active and reactive resistance of the converters into a DC signal. In this case, the semiconductor converter R_t is a current sensor, which is formed after the conversion of the active and reactive resistance of the converters.

It should be noted that humidity is one of the main quality indicators that affects the cost, physical and technological properties of substances and, first of all, such as grain and its processed products, wood and its derivatives, soil, mining products, food products, minerals fertilizers. In practice, it is necessary to measure humidity in the range from micro- and macro concentration (0,001÷0,1 %) to maximum saturation (80÷90 %) [11-15].

Let's consider some existing capacitive converters and their measuring circuits.

In a semiconductor-resistive humidity converter, the current-voltage characteristic of a semiconductor transistor is used. In the meter, a transistor and

measuring electrodes 1 and 2 are connected in series to the power source (Figure 1).

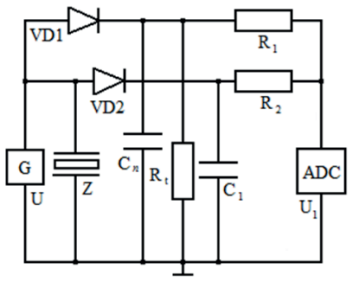


Figure 1: Semiconductor-resistive humidity converter.

The electric current flowing through these electrodes depends on the resistance value of the sample connected between them; the emitter of the transistor is connected to electrode 2 [16-18].

Figure 2 shows a functional diagram of a device for measuring the moisture content of bulk material in a process flow.

The flow of bulk material is divided into two parts, which fill capacitor sensors 2 and 3. Capacitor sensor 2 is connected to an electrical meter [19-20].

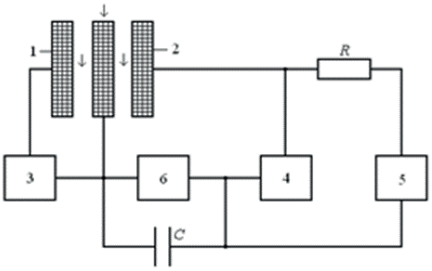


Figure 2: Functional diagram for measuring the moisture content of bulk materials in a process flow.

The flow of bulk material is divided into two parts, which fill capacitor sensors 2 and 3. Capacitor sensor 2 is connected to the electrical capacitance meter 4, capacitor sensor 3 is connected to the electrical capacitance meter 5 through a separating capacitor 6. When a high DC voltage is applied (2-3 K_v) an electric field is excited on the plates of the capacitor 3 sensor in its interelectrode space [21]. The compaction of bulk material in condenser 3 increases with increasing humidity. A capacitance ratio meter 6 is connected to electric capacitance meters 4 and 5. The disadvantage of the above capacitive sensor is that to excite an electric field in the interelectrode space, a high voltage is applied to the plates of the capacitor sensor, and also a change in the density of the bulk material leads to a decrease in the accuracy characteristics of the sensor.

In [22-25], a device for measuring humidity with a sensor consisting of two parts: reference and parametric was proposed. However, due to the peculiarity of the design of its sensor, such a device does not allow determining the soil parameters characteristic of it in a single localized volume, since the sensor, consisting of two parts, in accordance with its design, when measuring, is alternately connected to two adjacent volumes in the places where it is located. In addition, this device, which determines soil salinity using its frequency corrections, allows for a significant measurement error.

To determine the distribution of humidity in dispersed media, an electrical sensor consisting of a set of cylindrical electrodes is used. The disadvantage of this sensor is that the sensor is directly connected to the oscillatory circuit of the generator, i.e. The sensor is an integral element of the oscillatory circuit of the generator and changes in the capacitance of the sensor will lead to disruption of the entire oscillatory circuit, which results in measurement errors [26-27].

Figure 3 shows a block diagram of a device for measuring the humidity of dispersed media.

The test voltage from the measuring electrodes 3 and the comparative voltage from the reference capacitance are supplied to two identical amplitude detectors 5 and 6, and ensure the sensitivity of the device to the sign of the bridge mismatch. At the output of the amplitude amplifier, a difference signal is obtained, which is counted by the indicator.

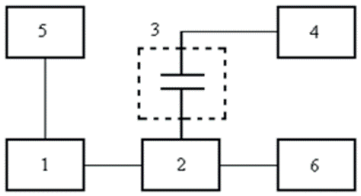


Figure 3: Block diagram of a device for measuring humidity dispersed media.

However, existing designs of converters for monitoring and control systems do not yet meet the requirements, because improvement and creation of new models of capacitive converters require the use of modern scientific achievements in the field of circuit solutions, the effective use of measuring circuits, and appropriate analysis methods.

Using the modified $Z-$transform of Laplace [28], the expressions of the instantaneous value of the output current I_n are obtained, according to which the graph is plotted (Figure 1).

The constant component of this current is $R_1 = R_2 = R$, when expressed by the (3):

$$I_1 = \frac{URR_t(R+2R_n)f}{(R+R_n)^2}\left[\left(e^{\frac{(R+R_n)\upsilon}{2R(R+2R_n)fC_n}}-1\right)e^{\frac{-(R+R_n)}{2R(R+2R_n)fC_n}}C_n - \left(e^{\frac{(R+R_n)\upsilon}{2R(R+R_n)fC_1}}-1\right)e^{\frac{-(R+R_n)}{2R(R+R_n)fC}}\right] \quad (3)$$

and the static characteristic of the capacitive-semiconductor humidity Converter of dispersed media is as follows (4):

$$U_1 = I_1 R_n, \quad (4)$$

where $f = 1/T_0 -$ is the frequency of the power supply.

In continuous feeding ($\upsilon = 1$) (3) is defined by the (5):

$$I_1 = \frac{URR_t(R+2R_n)f}{(R+R_n)^2}\left[\left(1-e^{\frac{-(R+R_n)}{2R(R+2R_n)fC_n}}\right)C_n - \left(1-e^{\frac{-(R+R_n)}{2R(R+2R_n)fC_1}}\right)C_1\right]. \quad (5)$$

and, therefore, the static characteristic of the capacitive-semiconductor humidity Converter is determined by the (6):

$$U_1 = \frac{URR_nR_t(R+2R_n)f}{(R+R_n)^2}\left[\left(1-e^{\frac{-(R+R_n)}{2R(R+2R_n)fC_n}}\right)C_n - \left(1-e^{\frac{-(R+R_n)}{2R(R+2R_n)fC_1}}\right)C_1\right]. \quad (6)$$

Expressions (3) and (4) show that, in General, the characteristic of the circuit transformation is nonlinear both in continuous and in pulsed power. For Figure 1, presents a family of static characteristics of the circuit transformation

$$U_1 = I_1 R_n = F(\Delta C)$$

and

$$U_1 = I_1 R_t = F(\Delta R),$$

constructed by the formula (4) for different values of the signal duration for the case of a differential converter, when

$$C_n = C_0 + \Delta C,\ C_1 = C_0 - \Delta C$$

and

$$R_t = R_0 + \Delta R,\ R_1 = R_0 - \Delta R.$$

A characteristic feature of the graphs is that at small values υ with growth ΔC and ΔR they deviate from the line in the direction of decreasing sensitivity, and for large $\upsilon -$ in the direction of increasing sensitivity. As a result, for each family of static characteristics of the capacitive-semiconductor converter, there is a certain value of the pulse duration υ, at which the characteristic of the circuit conversion is almost linear [29-35].

The Table 1 shows the main characteristics of the conversion circuit at the values of its parameters: $R = 47.10^3\,Om$, $\Delta C = 5.10^{-12}\,F$, $f = 25.10^4\,Hz$, $U = 30\,V$, $R_n = 10^6\,Om$.

The nonlinearity of the characteristics of the conversion of the input value of the capacitive-semiconductor humidity Converter of dispersed media is calculated as (7):

$$\beta = \frac{\Delta C_0}{S\Delta C_n} = \frac{U_{on} - S\Delta C_n}{S\Delta C_n}, \quad (7)$$

where ΔC_n and U_{0n}-limit values ΔC and U_{0n} for the range of moisture measurement of dispersed media; $S -$ sensitivity corresponding to the approximating line.

The sensitivity of the capacitive-semiconductor moisture Converter of dispersed media is calculated by the method of least squares according to the (8):

$$S = \frac{\sum_i U_{0i}\,\Delta C_i}{\sum \Delta C_i}. \quad (8)$$

Table 1: Key performance indicators of the transformation.

υ	U_{0n}	S, V/pF	ΔU_n, mV	β, %
0,25	1,00	0,21	-29,4	2,88
0,35	1,84	0,35	-29,8	2,71
0,4	2,16	0,32	-30,4	2,75
0,45	2,19	0,38	-32,6	2,82
0,5	2,25	0,46	-34,4	1,72
0,55	2,48	0,55	-32,6	1,61
0,6	3,15	0,64	-30,5	1,25
0,75	3,59	0,72	-5,41	0,15
0,8	3,83	0,76	6,7	0,17
0,85	4,04	0,81	19,0	0,48
0,9	4,22	0,84	31,1	0,74
0,95	4,33	0,86	40,4	0,94
1,0	4,38	0,88	44,1	1,02

Thus, tabular data indicate that for small pulse durations $\upsilon \le 0{,}75$, the absolute error of nonlinearity ΔU_n is negative and if $\upsilon \ge 0{,}75$-positive. At this boundary value υ, the nonlinearity β has a minimum value of only 0,15%. Comparison of data for continuous ($\upsilon = 1$) and pulsed power ($\upsilon = 0{,}75$) showed that the transition from continuous to pulsed power nonlinear characteristics of the conversion is reduced by more than six times. This fact gives practical value to the impulse power supply of the circuit, providing additional opportunities for linearization of its transformation characteristics.

4 CONCLUSIONS

Based on the above studies on the development of effective measuring circuits for capacitive-semiconductor measuring converters for measuring the humidity of dispersed media, the following conclusions were made:

1) Analysis of existing literature sources on the development of measuring circuits for capacitive-semiconductor measuring converters indicates that until now there are no effective measuring circuits for capacitive-semiconductor measuring converters that have high metrological characteristics. They mainly use alternating current with sinusoidal and quadrangular voltage as the source of the measuring circuit.
2) Mathematical modeling of the conversion of initial humidity based on inter-circuit graph transitions shows that depending on the type of capacitive-semiconductor measuring converter and the geometric parameters of the measured medium, there may be different distribution laws for the specific reactive and active conductivity of the converter and is described by the circuit function in the form of inter-circuit coefficients. They reveal the features of the distribution of specific reactive conductivity, which depends on the humidity of the dispersed medium.
3) Based on the above experimental studies, it has been established that at low and medium values (2÷6%) of the humidity of dispersed media, the bridge measuring circuit does not provide the required sensitivity value of the output signal of the capacitive-semiconductor measuring transducer. In this case, the output voltage, depending on the humidity of the measured environment, is reduced to 2÷4 mV with a simultaneous decrease in the sensitivity of the converter, therefore, additional measurement errors arise.
4) Pulse power supply of the measuring circuit of the capacitive-semiconductor measuring transducer with a semiconductor current sensor Rt ensures the linearity of the static characteristic. At the same time, measurement errors are significantly reduced as a result of linearization of the static characteristics of the converter. It has been established that the humidity conversion error is significantly related to the periodic charge and discharge of the capacitive-semiconductor measuring converter Cn, since when the charge and discharge of the converter are the same in time, the characteristic becomes linear, therefore, the measurement error decreases many times.
5) Thus, the results obtained indicate that at short pulse $\upsilon \leq 0.75$ durations the absolute nonlinearity ΔUn error is negative, $\upsilon \geq 0.75$ and at - positive. At this boundary value, υ the nonlinearity of the static characteristic β has a minimum value, only 0.15. A comparison of data for continuous ($\upsilon=1$) and pulsed power ($\upsilon=0.75$) showed that when switching from continuous to pulsed power, the nonlinearity of the conversion characteristic decreases by more than six times. This circumstance gives practical value to pulsed power supply of the circuit, which provides additional opportunities for linearization of its conversion characteristics.

REFERENCES

[1] A.V. Delgado, F. González-Caballero, R.J. Hunter, L.K. Koopal, and J. Lyklema. "Measurement and interpretation of electrokinetic phenomena." Journal of Colloid and Interface Science, vol. 309, 2007, pp. 194-224, [Online]. Available: https://doi.org/10.1016/j.jcis.2006.12.075.

[2] F. Zikrillayev, E.B. Saitov, J.B. Toshov, B.K. Ilyasov, and M.B. Zubaydullayev. "A Software Package for Determining the Optimal Composition and Parameters of a Combined Autonomous Power Supply System Based on Renewable Energy Sources." In AIP Conference Proceedings, 2022, vol. 2432, p. 020021.

[3] E.B. Saitov, "Optimal model for additional operation of the storage system for photovoltaic wind power plants." E3S Web of Conferences, Sustainable Energy Systems: Innovative Perspectives, SES 2020, vol. 220, 29 October 2020, [Online]. Available: https://doi.org/10.1051/e3sconf/202022001080.

[4] E.B. Saitov, Y.B. Sobirov, I.A. Yuldoshev, I.R. Jurayev, and S. Kodirov, "Study of Solar Radiation and Wind Characteristics in Various Regions of Uzbekistan." E3S Web of Conferences, Sustainable Energy Systems: Innovative Perspectives, SES 2020, vol. 220, 29 October 2020, [Online]. Available: https://doi.org/10.1051/e3sconf/202022001061.

[5] E.B. Saitov, "Renewable Energy Development in Uzbekistan: Current Status, Problems and Solutions." RSES 2020, E3S Web of Conferences 216, p. 01134, 2020, [Online]. Available: https://doi.org/10.1051/e3sconf/202021601134.

[6] I. Sapaev, E. Saitov, N. Zoxidov, and B. Kamanov, "Matlab-model of a solar photovoltaic station integrated with a local electrical network." IOP Conference Series: Materials Science and Engineering, 883, 1, 012116, IOP Publishing, 2020/7/1, [Online]. Available: https://doi.org/10.1088/1757-899X/883/1/012116.

[7] E.B. Saitov, Sh. Kodirov, B.M. Kamanov, N. Imomkulov, and I. Kudenov, "Increasing the

efficiency of autonomous solar photovoltaic installations for power supply of agricultural consumers." AIP Conference Proceedings 2432, p. 040036, 2022, [Online]. Available: https://doi.org/10.1063/5.0090439.

[8] F. Zikrillayev, E.B. Saitov, J.B. Toshov, B.K. Ilyasov, and M.B. Zubaydullayev, "A software package for determining the optimal composition and parameters of a combined autonomous power supply system based on renewable energy sources." AIP Conference Proceedings 2432, p. 020021, 2022.

[9] A.T. Rakhmanov and S.F. Kuznetsov, "Reliability of Semiconductor Converters of Temperature and Humidity of Disperse Media with Respect to Catastrophic Failures." International Journal of Advanced Research in Science, Engineering and Technology, vol. 8(12), December 2021, ISSN: 2350-0328.

[10] A.T. Rakhmanov, S.F. Kuznetsov, A.K. Khamrayev, and S.S. Gulyamov, "Matching of line-approximated static characteristics of temperature and moisture converters and automatic bridge." Science and World. International scientific journal, no. 4 (68), 2019, vol. 1, pp. 63-67.

[11] A.T. Rakhmanov, S.F. Kuznetsov, and A.K. Khamrayev, "Error of static characteristics of disperse media moisture when modeling the exponential process." Science and World. International Scientific Journal, vol. 1, no. 68, 2019, pp. 67-71.

[12] A.T. Rakhmanov and S.F. Kuznetsov, "Determination of Dynamic Errors of Converter Temperature and Humidity Measured Quickly Changing Output Signals." International Journal of Advanced in Science, Engineering and Technology, vol. 6, no. 4, 2019, pp. 8945-8949.

[13] E.B. Saitov and U.M. Akhmedov, "Tracking maximum power point of photovoltaic modules." AIP Conference Proceedings, vol. 2789, 2023, p. 060005.

[14] N. Zikrillayev, K. Ayupov, Z. Mamarajabova, E. Saitov, and I. Sabitova, "Adaptation and optimization of a photovoltaic system for a country house." E3S Web of Conferences, vol. 383, 2023, p. 04054.

[15] E. Saitov, I. Kudenov, F. Qodirova, D. Askarov, and M. Sultonova, "Analysis of the performance and economic parameters of different types of solar panels taking into account degradation processes." E3S Web of Conferences, vol. 383, 2023, p. 04059.

[16] E. Saitov, G. Khushakov, U. Masharipova, O. Mamasaliyev, and S. Rasulova, "Investigation of the working condition of large power solar panel cleaning device." E3S Web of Conferences, vol. 383, 2023, p. 04060.

[17] N. Zikrillayev, E. Saitov, J. Toshov, B. Muradov, and D. Muxtorov, "Autonomous Solar Power Plant for Individual use Simulation in LTspice Software Package Booster Voltage Converter." Proceedings of International Conference on Applied Innovation in IT, vol. 11, no. 1, 2023, pp. 207-211.

[18] B. Yuldoshov, E. Saitov, J. Khaliyarov, S. Toshpulatov, and F. Kholmurzayeva, "Effect of Temperature on Electrical Parameters of Photovoltaic Module." Proceedings of International Conference on Applied Innovation in IT, vol. 11, no. 1, 2023, pp. 291-295.

[19] E. Saitov, O. Jurayev, S. Axrorova, J. Ismailov, and B. Baymirzaev, "Conversion and use of Solar Energy Calculation Methodology for Photovoltaic Systems." Proceedings of International Conference on Applied Innovation in IT, vol. 11, no. 1, 2023, pp. 227-232.

[20] E.B. Saitov and T.B. Sodiqov, "Modeling an Autonomous Photovoltaic System in the Matlab Simulink Software Environment." AIP Conference Proceedings, vol. 2432, 2022, p. 020022.

[21] E.B. Saitov, Sh. Kodirov, B.M. Kamanov, N. Imomkulov, and I. Kudenov, "Increasing the Efficiency of Autonomous Solar Photovoltaic Installations for Power Supply of Agricultural Consumers." AIP Conference Proceedings, vol. 2432, 2022, p. 040036.

[22] E.B. Saitov, Y.M. Kurbonov, and B.M. Botirov, "Analysis of the influence of temperature on the operating mode of a photovoltaic solar station." IOP Conference Series: Earth and Environmental Science, vol. 614(1), 2020.

[23] E.B. Saitov, "Optimal model for additional operation of the storage system for photovoltaic wind power plants." E3S Web of Conferences, vol. 220, 2020.

[24] E.B. Saitov, Y.B. Sobirov, I.A. Yuldoshev, I.R. Jurayev, and S. Kodirov, "Study of Solar Radiation and Wind Characteristics in Various Regions of Uzbekistan." E3S Web of Conferences, vol. 220, 2020.

[25] E.B. Saitov, J.B. Toshov, A.O. Pulatov, B.M. Botirov, and Y.M. Kurbanov, "Networked interactive solar panels over the roof photovoltaic system (PVS) and its cost analysis at Tashkent state technical University." E3S Web of Conferences, vol. 216, 2020.

[26] E.B. Saitov, "Renewable energy development in Uzbekistan: Current status, problems and solutions." E3S Web of Conferences, vol. 216, 2020.

[27] E. Saitov and J. Toshov, "Portable autonomous solar power plant for individual use." E3S Web of Conferences, vol. 139, 2019.

[28] E.B. Saitov, M.K. Bakhadyrkhanov, S.A. Valiev, N.F. Zikrillaev, and S.A. Tachilin,"Silicon photovoltaic cells with clusters of nickel atoms." Applied Solar Energy (English translation of Geliotekhnika), vol. 52, no. 4, 2016, pp. 278-281.

[29] H. Schaumberg, "Sensors: Transl. from German/Under the general editorship," Tashkent: Tashkent State Technical University, 2002, 147 p.

[30] T.D. Japaridze, "Patent No. 171809 of the Russian Federation. Method for measuring the moisture content of bulk materials." 2002, no. 9, p. 164.

[31] Yu.M. Romanenko and V.A. Emelyanov, "Patent No. 379864 of the Russian Federation. Device for measuring soil moisture." 2005, no. 20, p. 78.

[32] A.I. Moiseev, "Development of capacitive converters of bulk materials." State and prospects for the development of information technologies: - Penza: Perm State Technical University, 2005, pp. 38-42.

[33] B.M. Kulwicki, "Ceramic sensors and transducers." J. Phys. Chem. Solids, vol. 45, no. 10, 2002, p.1015.

[34] "Universal converter. Universal transducer: Application no. 19929264." Germany, Knoll Meinkard, no. 19929264.

[35] A.I. Kuchimov. "Electronics and circuit design: Textbook." M.: Helios ARV, 2004, 336 p.

Development and Substantiation of Energy-Saving Methods for Controlling the Modes of Operation of Centrifugal Pumping Units in Complicated Operating Conditions

Javokhir Toshov[1], Azamat Makhmudov[2], Oybek Kurbonov[2], Golib Arzikulov[1,3] and Gulnoza Makhmudova[1]

[1]*Tashkent State Technical University named after Islam Karimov, University Str. 2, Tashkent, Uzbekistan*
[2]*Navoi State University of Mining and Technology, Tarobiy Str. 76V, Navoi, Uzbekistan*
[3]*University of Tashkent for Applied Sciences, Gavhar Str. 1, Tashkent, Uzbekistan*
javokhir.toshov@yandex.ru, azamat.maxmudov2015@yandex.ru, Oybek7001@mail.ru, miss.gulnoz@gmail.com

Keywords: Pump, Efficiency, Energy Efficiency, Operating Mode, Pressure, Flow, Throttle, Valve, Characteristic, Pressure, Fluid Flow.

Abstract: Rational and energy efficient operation of pumps for pumping and transporting liquids used in difficult conditions of mining is possible due to the implementation and use of modern methods of managing the operating modes of the pumps. Nowadays, the methods and resources used to regulate the operating modes of pumping units do not meet modern requirements, and many mines still use outdated methods of regulating the operating modes of pumps, which entails significant energy losses and an increase in the cost of the transported liquid. In the article, the methods of reducing electricity losses and the impact on the performance of pumping units were considered. It is concluded that this method of regulating work makes it possible to operate pumping units in the working area. Our article considers and experimentally substantiates the possibility of achieving savings in electrical energy, as well as increasing the reliability of the operation of pumping units based on the use of modern methods of regulating operating modes.

1 INTRODUCTION

Polymer One of the most effective ways to improve the performance and efficiency of pumps operating with a variable load and a sharp change in the content of the transported liquid, perhaps, is the use of adjustable electric drive. The performed studies of the results of the use of the adjustable electric drive showed that in some cases its installation leads to tangible energy savings, and in others it is insignificant, and thirdly, the installation of the drive does not provide economical operation. To study of the methods and forms of application of an adjustable drive indicates that in practice the technically simplest, but economically least effective ways of controlling pumping units, such as stabilizing the pressure at the pump outlet, are most often used. The degree of use of the energy saving potential, in this case, is no more than 16-26%, which leads to the fact that most of the potential, even after the installation of an adjustable drive, remains unclaimed. One of the main reasons for this situation is the insufficient knowledge of the influence of an adjustable electric drive on the operation of pumping units for pumping and transporting liquid [1].

However, the issues of choosing the optimal operating modes of pumps depending on the operating conditions have not been thoroughly studied. The study and determination of the optimal operating modes of pumping units under various operating conditions is an urgent scientific and practical problem, the solution of which leads to an increase in the reliability and efficiency of pumping units of industrial enterprises [2].

The aim of the research is to develop effective methods of energy saving in the operation of pumps based on the use of modern automation tools and optimization control methods.

The problem of the growing shortage of electricity can be solved most effectively on the basis of the development of energy saving systems and means. The reduction of energy costs is primarily achieved by increasing the efficiency of its use.

Pumping equipment of various technological cycles is one of the most significant consumers of electricity in industry [3].

The issues of energy saving during the operation of pumping units in complicated operating conditions is very relevant and requires increased attention from both designing and operating organizations.

2 METHODS

Operation of the pump in rational operating modes ensures its reliability, while operation in irrational modes is characterized by a decrease in efficiency [4].

In case of irrational operation, the following mechanical problems arise: failure of bearings, failure of mechanical seals, shaft failure and increased vibration [5].

The most common causes of equipment failure and the corresponding pump reliability curve depending on the location of the operating point are shown in Figure 1.

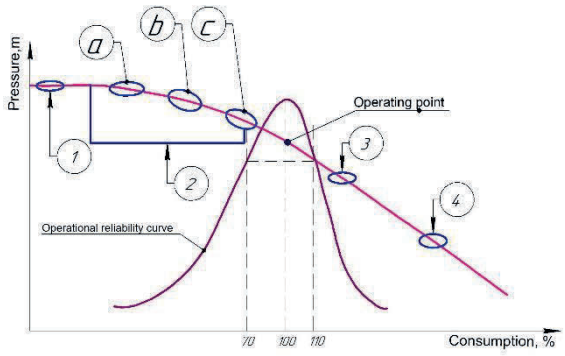


Figure 1: The reliability curve of the pump depending on the position of the working point: 1) a significant increase in temperature 2) a decrease in the service life of bearings and seals: a) due to vibration due to: possible cavitation, b) and c) the occurrence of flow recirculation at the inlet and outlet of the impeller, 3) a decrease in the service life of bearings and seals due to vibration caused by the separation of the flow in the flow part, 4) cavitation, overload of the electric motor.

Due to the occurrence of these problems, consumers often have a false opinion about the unreliability and inefficiency of pumps, while it is overlooked that the cause of the problems is incorrect operation in irrational modes. One of the main conditions for efficient and reliable operation of the pump is to find the operating point within its optimal operating range. The implementation of energy-saving measures determines the need to improve the performance of centrifugal pumps, which are the most common type of pumping equipment [6]. The basis for increasing the efficiency of centrifugal pumps is the improvement of operating conditions depending on the operating conditions, which significantly affect the efficiency of the centrifugal pump [7, 8].

In order to solve the problem associated with exceeded pump head and supply, it becomes necessary to change the operating parameters, which causes a change in the characteristics of the pump and system. The lack of regulation of pumping units significantly affects its efficiency. Since the volume of the transported liquid at enterprises is not constant and with small needs, part of the energy is lost [9, 10].

Choosing the optimal operating modes of a pump is a difficult task that determines the effectiveness of the main indicators of the installation as a whole, determining the optimal operating modes of a pump requires experimental research.

The issues of energy-saving modes and methods of effective control of pumping units have been considered by many authors who have studied the problems of optimization and improvement of operational parameters, while many dependencies affecting the operation of pumping units, such as changes in static and dynamic fluid levels, as well as external factors and fluid composition, have not been investigated [8-10].

To determine the main dependencies affecting the efficient operation of pumping units, we conducted experimental studies of the operating modes of the installation. The study of the performance characteristics of a centrifugal pump with various control methods was carried out on a laboratory installation SGU-SNS-012-6LR-PC, which contains a centrifugal pump, a pressure meter, a flow meter, controls and controllers. The general view of the laboratory installation SGU-SNS-012-6LR-PC, on which the experimental work was carried out, is shown in Figure 2.

During the experiments, the pressure-flow characteristics of the pump and the pumping station were measured, and the pressure and flow rate of the liquid were determined at various points in the system. During controlling the automation system of the stand with a PC, it is possible to implement a program-information action (PID) of regulation. For the toggle switch on the electronic control unit selects the automatic type of control, the control signal to the control device comes from the PC through the digital-to-analog conversion board (DAC).

To select the pressure and flow process variable PV to control the operation. It is possible to regulate

according to the following variables: pressure in front of the consumer, flow H, total flow.

Once a process variable has been selected, the electronic control unit screens display its current value. Next, enter the set value SP, to which the value of the process variable should come as a result of automatic control. The current error value e=SP-PV is displayed on the screen.

The current error value e=SP-PV is displayed on the screen. As a result of regression studies of the results, a mathematical model of calculation was established, in which the output signal at the i-th control step is generated by the program according to the following dependence (1):

$$u_i = K_P \cdot e_i + K_E \cdot \frac{e_i - e_{i-1}}{\Delta t} + K_U \cdot \sum e_i \cdot \Delta t, \quad (1)$$

where: Δt -is the time step between the control points.

The minimum time step is 0.4 s, if you set a smaller value, the real time step will be equal to the specified one.

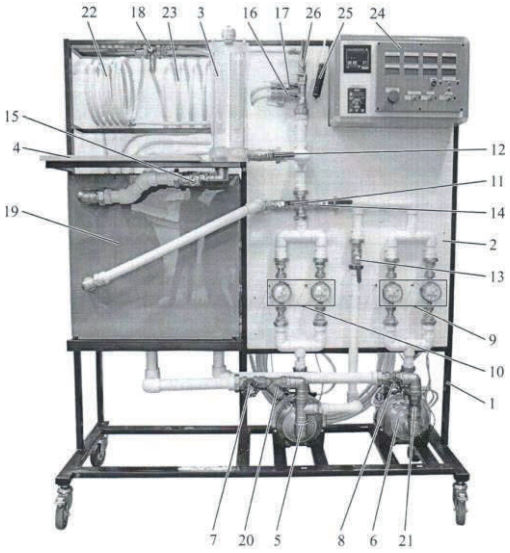


Figure 2: Laboratory stand SSU-CNS-012-6LR-PC: 1) frame, 2) mounting plate, 3) dimensional container, 4) shelf for PC (laptop), 5) centrifugal pump with frequency control, 6) centrifugal pump, 7), 8) valves, 9), 10) flow meters, 11), 12), 13), 14), 15) ball valves, 16), 17), 18), 26) taps, 19) tank, 20), 21) filters, 22), 23) pipelines, 24) electronic control unit, 25) stopwatch control knob.

This minimum step is determined by the time of input of information from the sensors through the analog-to-digital conversion board (ADC) in the PC. The step should be set in the text field of the screen. Factors of performance (KP), changes in engine power input (KE), changes in utilization (KU) should be set in the corresponding text fields of the zone on the screen. The range of the output signal and is 0-10 V. The dimension of the coefficients KP, KE, KU is determined by the choice of the process variable.

3 RESULTS AND DISCUSSION

During the experiments, the pressure and flow characteristics of the pump and pumping station were measured, the pressure and flow rate of the liquid at various points of the system were determined.

The results of the experimental studies made it possible to obtain the following results, which will lay the foundation for energy saving during the operation of pumping units in various conditions for the quality of the transported liquid.

Experimental work allowed us to obtain the following results:

- the dependence of the pump power consumption (W) on the pump supply when regulating the supply using a proportional valve;
- the dependence of the pump power consumption (W) on the pump supply when regulating the supply using a frequency converter;
- the dependence of the pump power consumption (W) on the pump supply when regulating the supply using a frequency converter and a PC.

Figure 3 shows the graphical dependence of the pump power consumption (W) on the pump supply (l/min) when regulating the supply using a proportional valve. On which it is observed that with an increase in the pump supply for every 14 l/min, the power consumption increases by 150 watts.

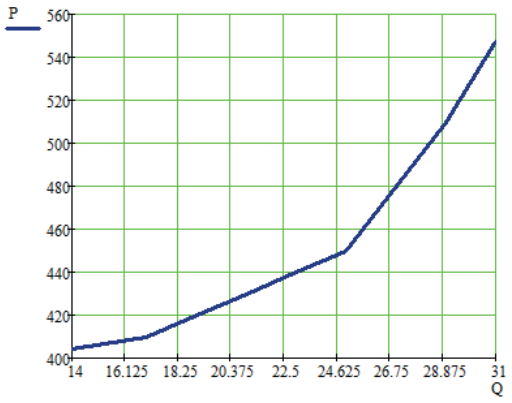


Figure 3: Graphical dependence of the pump power consumption (P, W) on the pump supply (Q, l/min) when regulating the supply using a proportional valve.

Figure 4 shows the graphical dependence of the pump power consumption (W) on the pump supply (l/min) when regulating the supply using a frequency converter.

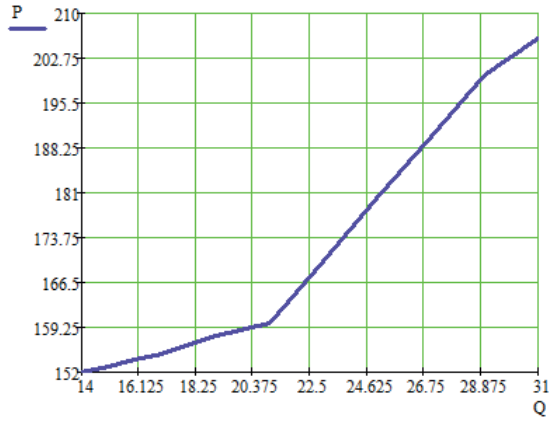


Figure 4: Graphical dependence of the pump power consumption (P, W) on the pump supply (Q, l/min) when regulating the supply using a frequency converter.

Figure 4 also shows an increase in power consumption, in this case, with an increase in the pump supply for every 14 l/min, there will be an increase in power consumption by 50 watts. Thus, the use of a frequency converter allows you to get the required supply, using only 60 watts of electrical energy.

The savings when using a frequency converter instead of a proportional gate valve is about 250 watts, for 14 l/ min, which is also presented in the form of a graph in Figure 5.

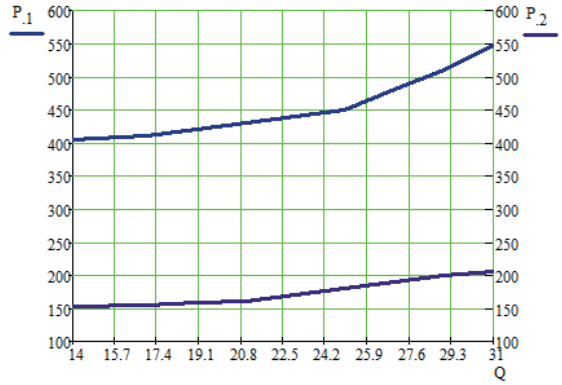


Figure 5: Graphical dependence of the pump power consumption (W) on the pump supply (Q, l/min) when regulating the supply using a proportional valve (P1, W) and a frequency converter (P2, W).

The most economic benefits of the operation of pumps can be obtained by using a frequency converter and PC when regulating operating modes. Figure 6 shows the graphical dependence of the pump power consumption (W) on the pump supply when regulating the supply using a frequency converter and a PC.

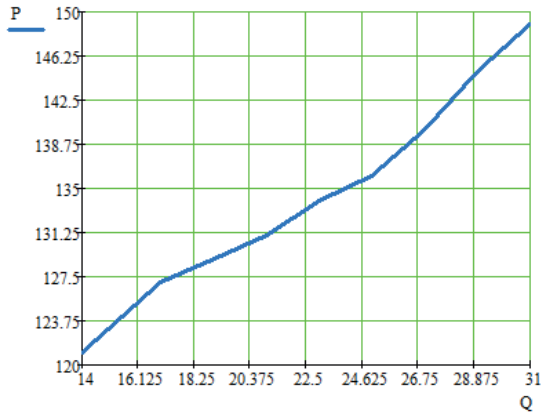


Figure 6: Graphical dependence of the pump power consumption (P, W) on the pump supply (Q, l/min) when regulating the supply using a frequency converter and a PC.

From Figure 6, it is observed that the use of a frequency converter and a PC to regulate the pump supply contributes to significant savings in electrical energy consumed by the pump drive, while the savings in electrical energy consumed for the same received supply is 110 watts.

The effectiveness of a particular method of regulation is determined by the characteristics of the system and its change over time. In each case, it is necessary to make a decision depending on the operating conditions and characteristics of the system. When choosing a method of regulation, it is very important not to fall under the influence of stereotypes that have developed recently. So, according to one of them, the use of a frequency converter helps to reduce energy consumption. However, frequency control does not always lead to a reduction in power consumption, and sometimes even has the opposite effect. The use of frequency control is advisable when pumps are operating in a system where friction losses prevail (friction losses in pipelines and shut-off valves [11].

CONCLUSIONS

The research carried out by us allowed us to obtain the following main conclusions, which in the future allow us to successfully choose the optimal methods

for regulating the operating modes of centrifugal pumping units. Due to the difference in water inflow, an unregulated motor gives the pumping unit a change in operating mode, which contributes to an increase in the occurrence of accidents in the operation of pumping units. The proposed control method makes it possible to control pumping units with optimal engine performance. Contributes to the operation of pumping units in a reliable operating area, in which the occurrence of reasons for stopping pumps is less than at other points of operation.

In the course of the conducted research, it was found that the use of the proposed equipment allows to reduce energy costs by 67% during the operation of pumps.

In addition, it helps to increase the efficiency of pumps, reduces the likelihood of sudden failures, increases the safety of workers.

REFERENCES

[1] V.I. Bessarab, R.V. Fedyun, and V.A. Popov, "Control of the mining drainage plant in emergency and abnormal modes of the work," Scientific works of DonNTU, Donetsk, vol.106, pp. 26-33, 2006.

[2] V.G. Mikhailov and P.V. Petrov, "Calculation and experimental method for diagnosing the degradation of some units of submersible centrifugal pumps," Bulletin of USATU, Ufa, vol. 1, pp. 82-93, 2009.

[3] Yu.P. Stashinov, D.A. Bochenkov, V.V. Volkov, "Energy-saving regulation of the operation mode of the main drainage installations of mines and pits," Notes of the Mining Institute, Moscow, vol. 192, pp. 209-211, 2011.

[4] J.B. Toshov, B.R. Toshov, B.N. Baratov, and A.L. Haqberdiyev, "Designing new generation drill bits with optimal axial eccentricity," Mining Informational and Analytical Bulletin, Moscow, vol. 9, pp. 133-142 2022, doi: 10.25018/0236_1493_2022_9_0_133.

[5] R. Ergashev, O. Azizov, N. Dehkanova, and A. Bozorov, "Development of energy-saving modes of irrigation pump stations," IOP Conf. Series: Materials Science and Engineering, vol. 883, pp.1-6, 2020, doi:10.1088/1757-899X/883/1/012017.

[6] I. Bakman, "High-Efficiency Predictive Control of Centrifugal Multi-Pump Stations with Variable-Speed Drives" PhD thesis, Tallinn University of Technology, Tallin, p. 149, 2016.

[7] Y. Wang, K. Luo, K. Wang, H. Liu, Y. Li, and X. He, "Research on pressure fluctuation characteristics of a centrifugal pump with guide vane" Journal of vibroengineering, Lithuania, vol. 19 (7), pp. 5482-5497, 2017.

[8] L. Gevorkov, "Simulation and Experimental Study on Energy Management of Circulating Centrifugal Pumping Plants with Variable Speed Drives" PhD thesis. Tallinn University of Technology, Tallin, p.115, 2017.

[9] I. Nikolenko and P. Shvagirev, "Analysis of methods of improving energy efficiency of pumping stations power units of water supply systems," Econtechmod. an International Quarterly Journal, no. 1, vol. 7, pp. 67-77, 2018.

[10] A. Makhmudov, O.M. Kurbonov, and M.D. Safarova, "Research of the pressure characteristics of the centrifugal water drainage plant of the WCP 25-60G brand," Australian Journal of Science and Technology, Melbourne, vol. 4(2), pp.279-282, June 2020.

[11] M. Bogdevicius, J. Januteniene, R. Didziokas, S. Razmas, V. Skrickij, and P. Bogdevicius, "Investigation of the hydrodynamic processes of a centrifugal pump in a geothermal system," Transport, Lithuania, vol. 33 (1), pp. 223-230, 2018.

Analysis of Electric and Thermal Efficiency of Crystal Silicon Small Power Suppliers

Muhammadjon Tursunov[1], Khabibullo Sabirov[1], Tohir Axtamov[1], Umirbek Abdiyev[2], Boysori Yuldoshov[2], Jasur Khaliyarov[2], Sardor Bobomuratov[2] and Sirojiddin Toshpulatov[2]
[1]*Physical-Technical Institute, Chingiz Aytmatov Str. 2B, Tashkent, Uzbekistan*
[2]*Termez State University, Barkamol Avlod Str. 43, Termez, Uzbekistan*
muhammadtursunov54@gmail.com, sabirovhabibullo@gmail.com, axtamovtz@uzsci.net, umr79@mail.ru, b.yuldoshov10@mail.ru, xjxjasur@mail.ru, bobomuratovsardor1@gmail.com, toshpulatovs@tersu.uz

Keywords: Solar Radiation Intensity, Short Circuit Current, Open Circuit Voltage, Electric Power.

Abstract: The article presents the results of studying the electric and thermal efficiency of mobile photovoltaic and photothermal devices (MPVD and MPTD) based on crystalline silicon with a power of 60W panel, designed for the production of electricity and hot water for the domestic needs of rural residents. Our experiment conducted in natural conditions was carried out in the heliopolygon of the Physical-Technical Institute in July of the summer season. The device was manually adjusted and measurements were carried out in the directed mode. Reflectors are installed on the side of the photovoltaic panel (PV) in order to increase the intensity of sunlight in MPTD. As a cooling system, heat collector (HC) is installed in the back of the PV. The HC of the photothermal panel (PVT) is fundamentally different from the system of HC in other works. During the experiment carried out in natural conditions, the effects of the reflectors installed on the front surface of the photothermal device's PV and the heat collector attached to the back side were determined. Thermal efficiency was 5.5-6 times higher than electric efficiency for PVT panels. And 90% of the solar energy falling on the surface of the photothermal device is converted into useful energy.

1 INTRODUCTION

It is well known that the amount and volume of natural energy sources is declining for years. This demand for the rational use of renewable energy sources is an important task for humanity. Taking into account that the share of renewable energy sources will increase in the future, there is a lot of research that needs to be done by scientists in this field. The use of wise from solar energy is a key role in the development of work in this area. Several important factors adversely affect electricity in the process of obtaining power using a PV panel. These examples, the PV panel can be said high temperatures, pollination, and effectiveness of system devices. PV panel is a system consisting of solar cells based on the p-n junction, which converts light energy from the sun into electrical energy. In this case, the part of the energy that is not converted into electricity, but is absorbed by the solar cells, causes the PV to heat up. It is well known that the PV panel efficiency falls with an increase in its temperature [1-4]. Therefore, the PV temperature has practical and theoretical work on cooling, through which the goal is to effectively use thermal energy. The active, passive, and hybrid types of cooling PV are among them, including these methods differ from each other in the use of purpose and function [5-7]. A system of PV panel combined with a heat collector is called a PVT panel. Different types of heat collectors are used to cool PVT panels. As a coolant, various liquids, mainly water, in some cases air are used and they are passed through heat collectors. Such systems, in turn, need to evaluate the system's thermal efficiency. Heat collectors of different materials and shapes were tested in PVT panels [8,9]. The effect of reflectors on PVT parameters is widely covered in works such as [10-12], and its main function is to increase the radiation energy coming from the sun and protect PV from dusting. At present, it is necessary to develop energy devices based on PV and PVT panel and apply them to the population in rural areas. Using the results

obtained on MPVD and MPTD, we assessed the electrical and thermal effects of these devices.

2 METHODS AND MATERIALS

The research is extensive and the main goal is to compare the parameters of two mobile devices. In this case, in our work, the electric and thermal efficiency of these two devices were calculated. One of the main goals of our work is to influence the effects of these reflectors and heat collectors on PV panel effects. Through this, the heat collector's thermal efficiency was calculated. Monocrystalline silicon PV was used in the device with a maximum power of 60W. The PV panel datasheet and PVT panel parameters based on it are presented in Table 1.

Table 1: Geometric dimensions, physical and technical characteristics of PVT parts.

Parameters	Dimension
Geometric dimensions	
PV surface, S_{PV}	0,36m^2
PV frame width, d	2.5sm
Reflector surface, S_{ref}	0,36m^2
Reflector thickness, d_{ref}	0,4sm
The surface of the back cover, S_q	0,36m^2
The thickness of the back cover, d_q	0,4sm
Physical and technical characteristics	
Maximum power of PV, P_{max}	60W
Electrical efficiency of PV, η	16,5%
Open circuit voltage of PV, U_{oc}	24,4V
Short circuit current of PV, I_{sc}	3,33A
The fill factor of the PV's volt-ampere characteristic, FF	0,74
Reflection coefficient of the reflector, R	0,5
Water capacity of the heat collector, V	9 litre

A heat collector and two reflectors along the long sides of the frontal surface are installed on the back of the PV (Figure 1).

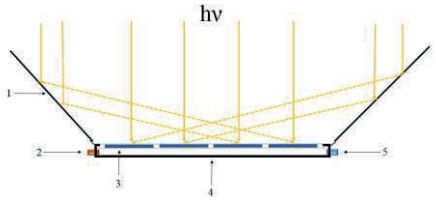


Figure 1: Structure of PVT. 1-reflector, 2-outlet water, 3-PVT, 4-back cover,5-inlet water.

The heat collector is specially designed. To achieve constant heat exchange with the back surface of the PV, liquid inlet and outlet devices are installed in the PV structure. The back side is closed based on alucobond material, which serves as a cover. An overview of the MPTD from the different cases is shown in Figure 2. Reflectors are made by glueing aluminium foil to an alucobond. The total surface of the reflectors is equal to the surface of the PV panel. They can be manually adjusted to the desired angle during the experiment. The advantage of the device is that it is easy to move the MPTD from one place to another, and it is easy to install it at the right angle.

Figure 2: General view of the MPVT from the different cases.

2.1 The Function of Reflectors and Collector in the Study

The main function of reflectors is to increase the energy of solar radiation falling on the PV surface. This is very important for winter, spring and autumn months due to increasing solar radiation. It is known that the PV parameters measured AM=1.5, E=1000W/m^2, T=25^{0}C are accepted as standard conditions. However, the amount of solar radiation specified in this standard is not maintained at almost all points on the earth's surface, and to ensure a situation close to this condition, additional devices are required, or the electrical parameters of PV corresponding to the value of solar radiation for that point are determined. In this article, the reflector serves to increase solar radiation and the heat collector serves to cool the PV. The energy falling on the PV surface in the reflector operating state is determined as (1):

$$W = W_{PV} + 2 \cdot W_{ref} . \quad (1)$$

It can be seen from this formula that additional energy was generated under the influence of the reflector. The task of the heat collector is to reduce PV electrical losses (at the expense of the standard 25 ºC excess of temperature), and if necessary, taking additional heat energy from PV. If the electrical losses are reduced by cooling the PV, by increasing the voltage and power, thermal energy can be explained by the fact that the temperature of the liquid leaving the heat collector is higher than the liquid entering the collector. This energy can be calculated as (2):

$$Q = m \cdot c \cdot \Delta t \ . \quad (2)$$

The system energy efficiency assessment is done as follows. Of course, electricity and thermal effects are calculated separately and added then.

2.2 Electric and Thermal Efficiency in Our Research

The total efficiency was calculated as (3):

$$\eta_{total} = \eta_{el} + \eta_{ther} \cdot \quad (3)$$

We can calculate the heat efficiency as (4):

$$\eta_{ther} = \frac{Q_{useful\_heat}}{S_{HC} \times E}. \quad (4)$$

The useful heat Q_{useful_heat} can be calculated from (5):

$$Q_{useful\_heat} = \dot{m} \cdot C_p \cdot \Delta T. \quad (5)$$

Where $\dot{m}$ is the fluid (air, water…etc, in our work this is water.) mass flow rate; C_p is the used fluid's specific heat, and ΔT is the fluid temperature difference between the inlet and outlet. The electrical efficiency is calculated in standard form (6):

$$\eta_{el} = \frac{FF \cdot I_{sc} \cdot U_{oc}}{S_{pv} \times E}, \quad (6)$$

where, FF is fill factor, I_{sc} – the short circuit current, U_{oc}– the open circuit voltage, E – the solar radiation intensity in [W/m^2], S_{pv}– the PV panel area.

3 RESULTS AND DISCUSSION

Our main purpose is to compare MPVD and MPTD electric parameters and study the electrical and thermal efficiency of the second device. In terms of reflectors and collector advantages in our second device, their effects are very important. The measurement results were measured on July 18, 2022, on an open sunny day at the heliopolygon of the Physical-Technical Institute. During the experiment, the temperature ranged from 34^{0}C to 41^{0}C, and the wind speed was 0.2–5m/s, and the air humidity was 14.2%. Our experiment was conducted in tracker mode (manual guidance mode). In order to ensure direct sunlight (90^0) to the front surface of the solar panels, the operator manually directed it at intervals of 15-20 minutes. MPVD and MPTD parameters were simultaneously measured for the reliability of the obtained data and parameter comparison. Figure 3 shows an overview of MPVD and MPTD in the measurement process. In this case, these two systems' electric and thermal parameters are also calculated and compared. The purpose of the work is to study the variation of PV and PVT short-circuit current (I_{sc}), open circuit voltage (U_{oc}) and produced electric power (P) under natural solar radiation conditions (PTI Heliopolygon). A different from MPVD, MPTD has reflectors on its side that increase the intensity of sunlight up to 1.5 times. Reflectors are made of 4 mm thick alukobond, and aluminum films with a reflection coefficient equal to 0.5 are glued to it as a coating. At night, the reflectors are used as a cover to protect the front surface of the PVT from dust and pollution.

The heat collector part of PVT panel is designed to increase cooling efficiency and reduce energy losses. Two water inlet and outlet valves are installed in the heat collector structure. The volume of the heat collector is 9 liters. MPVD and MPTD parameters were simultaneously measured for the reliability of the obtained data and parameter comparison.

Figure 3: An overview of MPVD and MPTD in the measurement process.

There are 3 modes for MPTD. 1 - measurements in PV mode: from 9:00 a.m. to 10:40 a.m. The current is directly proportional to the solar radiation incident on the PV and PVT surfaces, and in this mode, the solar radiation incident on the surface of both devices is the same. However, since the PVT surface has a higher temperature than the PV surface, there is a change in the values of U_{oc}. 2 - the reflectors are open: due to the increase in temperature due to the increase in the intensity of solar radiation, the value of U_{oc} of

PVT decreased to 20.45V and I_{sc} increased to 4.05A, and at the same time, the U_{oc} value indicator of PV increased to 21.4V, and I_{sc} to 3.05A which is equal. Since the back surface of PVT does not exchange convective heat with the atmosphere in a closed environment, the value U_{oc} is small, but due to the large radiation intensity on the PVT surface, I_{sc} is also relatively large. 3- in PVT mode, we can see that the value U_{oc} rises to 23.5V at PVT and I_{sc} remains constant around 4A due to the triggering of HC.

When the reflectors are placed at the optimal angle (mode 2), the short-circuit current increases from 3.05A to 4.1A as a result of the increase in radiation intensity from 880W/m^2 to 1260W/m^2. In mode 3, the short circuit current remained almost unchanged. During the experiment, the solar radiation intensity versus time graph is presented in Figure 4. We can see from the graph that the reflectors increase the intensity of sunlight up to 1.5 times. As a result of collector cooling, the open circuit voltage difference was restored to 2.3V, and an electrical recovery of 20% was achieved for this voltage.

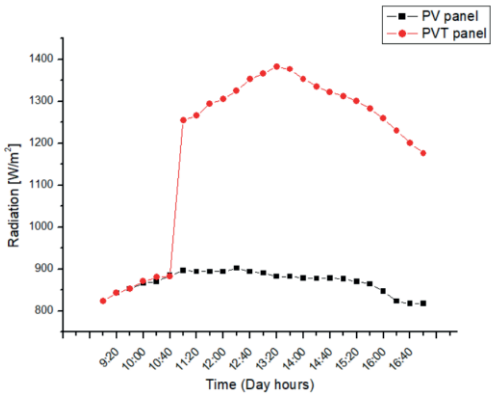


Figure 4: Time dependence of solar radiation.

The electrical power of the devices can be calculated using (7), based on the values of the parameters obtained during the measurements:

$$P = FF \cdot I_{sc} \cdot U_{oc} \cdot \qquad (7)$$

The time dependence graph of PV and PVT panel power are shown in Figure 5. It is clear from this graph that in normal cases, 60W PV capacity cannot produce 60W energy at any part of the day. In this case, the maximum power of the PV is around 50W, and considering that in our experience we measured the device in the directed mode, this result means that it will be smaller if the device is in a stationary state.

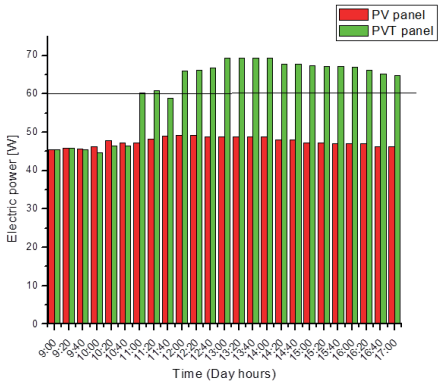


Figure 5: Time dependence of PV and PVT panel electric power.

The goal of the experiment is to maintain the maximum PV power for as long as possible during working hours. This purpose is achieved in the example of PVT. In our experiment, the inlet water temperature is not changed, and at the 3 different water flows (0.005, 0.01 va 0.015 kg/s) additional 1.5l, 3l va 4.5liters of hot water at per minute was obtained from the heat collector. This is proof that the heat losses can be used effectively. The calculations of the average capacity of the PV Panel was 47.6W, and the average power of the PVT panel was 61.6W. In the hot chilla of summer, our above experiment proved that it is possible to make good use of the fixed capacity of PV.

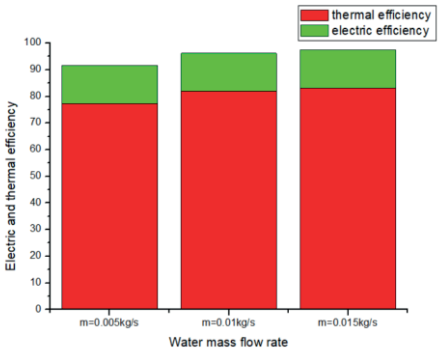


Figure 6: Electric and thermal efficiency of PVT panel for different water flow.

The collector's thermal efficiency was 77.33%, 81.88% and 83.24% respectively for three different water flows. PVT panel electric efficiency, in this case, was almost unchanged upper 14% and for PV panel electric efficiency was above 15%. However, it is to note that PVT's total efficiency was 90% -95%. We can see this clearly in Figure 6 above.

4 CONCLUSIONS

The electrical parameters of MPVD and MPTD used in our experiment were compared. It was found that 60W monocrystalline PV does not provide this rated power at any time of the day, due to insufficient solar radiation and overheating of the PV. In PVT panel, specified power was achieved to 69W by reducing electricity and heat losses. As well as the increase in the average power to 29% was determined by calculations. As a result of the experiment, the following conclusions were obtained:

- Thermal energy obtained from PVT are 5.5-6 times higher than electric energy.
- Reflectors should be used in low radiation days and heat collector systems.
- The energy, which falls on the PVT panel surface, can become 90% profitable energy.
- In PV Panel, it was found that this figure was around 15%.

If we cool PV panels effectively, it is possible to reduce their electrical losses, and also we can get additional thermal energy.

REFERENCES

[1] B. Yuldoshov, E. Saitov, J. Khaliyarov, S. Toshpulatov, and F. Kholmurzayeva, “Effect of Temperature on Electrical Parameters of Photovoltaic Module,” Proceedings of International Conference on Applied Innovation in IT, 2023, 11(1), pp. 291-295.

[2] K. Kant, A. Shukla, A. Sharma, and P. H. Biwole, “Thermal response of poly-crystalline silicon photovoltaic panels: numerical simulation and experimental study,”. Sol. Ener., vol. 134, 2016. pp. 147-155. [Online]. Available: https://doi.org/10.1016/j.solener.2016.05.002.

[3] N. Aoun, “Methodology for predicting the PV module temperature based on actual and estimated weather data,” Ener. Conv. and Man.: X, vol. 14, 2022. p. 100182. [Online]. Available: https://doi.org/10.1016/j.ecmx.2022.100182.

[4] A. Nahar, M. Hasanuzzaman, N. Rahim, and S. Parvin, “Numerical investigation on the effect of different parameters in enhancing heat transfer performance of photovoltaic thermal systems,” Ren. Ener., vol. 132, 2019. pp. 284–95. [Online]. Available: https://doi.org/10.1016/j.renene.2018.08.008.

[5] G. Yuanzhi, W. Dongxu, D. Zhaofeng, W. Changling, Z. Liutao, Z. Jili, X. Guoying, and Z. Xiaosong, “A passive evaporative cooling strategy to enhance the electricity production of hybrid PV-STEG system,” Applied Energy, vol. 349, 2023, [Online]. Available: https://doi.org/10.1016/j.apenergy.2023.121689.

[6] A.M. Elbreki, K. Sopian, A. Fazlizan, and A. Ibrahim, “An innovative technique of passive cooling PV module using lapping fins and planner reflector,” Case Studies in Thermal Engineering, vol. 19, 2020, [Online]. Available: https://doi.org/10.1016/j.csite.2020.100607.

[7] M. Abd-Elhady, Z. Serag, and H. Kandil, “An innovative solution to the overheating problem of PV panels,” Energy Convers. Manag., vol. 157, pp. 452-459, 2018.

[8] H. Li, P. Hou, and Y. Sun, “Photoelectricity and photothermal performance experiment on solar photovoltaic/loop-heat-pipe water heating system,” Trans. Chin. Soc. Agric. Eng., 34 (7) (2018), pp. 235-240.

[9] R.G. Mohamed, A. Mohsen, and R. Hegazy, “Temperature distribution modeling of PV and cooling water PV/T collectors through thin and thick cooling cross-fined channel box,” Energy Reports, vol. 8, 2022, pp. 1144-1153, [Online]. Available: https://doi.org/10.1016/j.egyr.2021.11.061.

[10] M. Asvad, M. Gorji, and A. Mahdavi, “Performance analysis of a solar module with different reflectors and cooling flow fields,” Applied Thermal Engineering, vol. 219, 2023, [Online]. Available: https://doi.org/10.1016/j.applthermaleng.2022.119469.

[11] P.V. Sai and K.S. Reddy, “Theoretical and experimental investigations of inclined solar still using secondary reflectors on east-west for enhanced distillate yield,” Applied Thermal Engineering, vol. 236, 2023, [Online]. Available: https://doi.org/10.1016/j.applthermaleng.2023.121836.

[12] L. Marzieh, A.Sh. Hossein, and F. Mohammad, “Experimental study on simultaneous use of phase change material and reflector to enhance the performance of photovoltaic modules,” Journal of Energy Storage, vol. 54, 2022, [Online]. Available: https://doi.org/10.1016/j.est.2022.105342.

APPENDIX

The list of abbreviations used in this paper is given below:

W	total light energy (J)
W_{PV}	light energy incident on the PV surface (J)
$W_{ref.}$	light energy returned from the reflector (J)
E	radiation intensity (W/m^2)
S	surface (m^2)
Q	amount of heat (J)
m	mass (kg)
c	specific heat capacity (J/kg·^{o}C)
Δt	temperature difference (^{o}C)
I_{sc}	short circuit current (A)
U_{oc}	open circuit voltage (V)
P	electric power (W)