Perioperative monitoring of facial nerve function in vestibular schwannoma surgery for objective determination of postoperative nerve function with machine learning approaches

Dissertation

zur Erlangung des akademischen Grades

Doktor der Medizin (Dr. med.)

vorgelegt

der Medizinischen Fakultät

der Martin-Luther-Universität Halle-Wittenberg

von Magdalena Holze

Betreuer*in: 1) Priv. Doz. Dr. med. Stefan Rampp

2) apl. Prof. Dr. med. Julian Prell

Gutachter*innen: 1) Prof. Dr. med. Kathleen Seidel, Bern

2) Priv. Doz. Dr. med. Carolin Weiß Lucas, Köln

Datum der Verteidigung: 09.01.2024

Report

In vestibular schwannoma surgery, the facial nerve and its course are of utmost relevance due to the fact that the nerve can potentially be damaged which then is associated with significant restriction in the quality of life. For functional preservation and to differentiate between already preoperatively existing nerval impairment and intraoperatively arising nerve deficit, it is important to determine the exact facial nerve function in the perioperative course by nerve monitoring methods. For example, intraoperative free-running EMG can be used to record A-trains, typical pathophysiological patterns which correlate in their total quantity - the traintime - with postoperative facial nerve function and can predict a postoperatively occurring facial nerve impairment. Traintime is influenced i.a. by the intermediate nerve. In the presence of such a split nerve, additional A-train activity may be triggered, which, however, does not correlate with any postoperative facial palsy. Another issue is the clinical evaluation and grading of pre- and postoperative facial nerve function itself. The established House-Brackmann facial nerve grading system (HBGS) is influenced by subjective impression, and introduces considerable rater-related variability. The aim of the presented studies was to obtain an objective estimate of facial nerve function using surface electromyography (sEMG) measurements for grading and neuronal network analysis including, among others, multiple data from intraoperative nerve monitoring in the perioperative vestibular schwannoma surgery setting. sEMG was used in a prospective exploratory study design with 28 patients. In total, 59 measurements were conducted in the pre- and postoperative period. Additionally, HB grades were determined. In this exploratory trial, multivariant machine learning analysis in different scenarios (differentiation of each, normal (1) and slight (2) vs. impaired facial nerve function and classification of HB 1-3 (3)) were performed. Overall, good differentiation capability of the HB estimated with machine learning compared to the clinical determined HB could be shown as well as the ability to distinguish clinically relevant differences. Accordingly, sEMG could be demonstrated to be a potential alternative method for facial nerve grading compared to the current HBGS. In the second study, data from 200 patients including continuous intraoperative EMG recordings, traintime and occurrence of A-train clusters (A-trains occurring in most of the recording channels within a short time) were retrospectively analyzed. Neural feed-forward networks were trained to estimate postoperative HB grades based on these multimodal data. Neuronal network analysis showed satisfying performance estimating postoperative HB grades. Especially networks based on the factors traintime, preoperative HB grade and tumor size achieved more precise results than networks based on tumor size or mean traintime alone. Based on these study data, the methodologies of sEMG and neural network (NN) analysis can be used in further studies to improve monitoring of facial function in real-time.

Holze, Magdalena: Perioperative monitoring of facial nerve function in vestibular schwannoma surgery for objective determination of postoperative nerve function with machine learning approaches, Halle (Saale), Univ., Med. Fak., Diss., 23 pages, 2023

Referat

Bei der operativen Resektion eines Vestibularschwannoms sind der N. facialis und sein Verlauf von großer Relevanz, da dieser potenziell geschädigt werden kann, was mit einer signifikanten Einschränkung der Lebensqualität verbunden ist. Für die funktionelle Erhaltung und zur Unterscheidung zwischen einer bereits präoperativ bestehenden Nervenschädigung und einem intraoperativ entstehenden Defizit ist es notwendig, die genaue Nervenfunktion im perioperativen Verlauf mittels Neuromonitoring zu bestimmen. So können z.B. mit dem intraoperativen freilaufenden EMG A-Trains aufgezeichnet werden, typische pathophysiologische Muster, deren Gesamtmenge (Traintime) mit der Fazialisfunktion korreliert und so das Auftreten einer postoperativen funktionellen Schädigung vorhergesagt werden kann. Diese A-Trains werden unter anderem durch den N. intermedius als Confounder beeinflusst. Insbesondere bei Verletzung eines separat verlaufenden N. intermedius wird zusätzliche A-Train-Aktivität ausgelöst, die jedoch nicht wesentlich mit der postoperativen Facialisfunktion korreliert. Zudem müssen weitere individuelle Faktoren beim Neuromonitoring berücksichtigt werden, die durch die gängigen, teils subjektiven, Graduierungssysteme wie das HBGS nicht vollständig erfasst werden. Ziel dieser Studien war es, perioperativ die Fazialisfunktion u.a. mit Oberflächen-EMG (sEMG) objektiv zu ermitteln. Die Messung des N. facialis mittels sEMG wurde in einem prospektiven, explorativen Studiendesign bei 28 Patienten eingesetzt und insgesamt 59 Messungen perioperativ durchgeführt sowie die klinischen HB-Grade erfasst. Es wurden multivariante Machine Learning Analysen zu verschiedenen Vergleichsszenarien durchgeführt: Differenzierung von (1) regelrechter vs. beeinträchtigter, (2) leicht eingeschränkter vs. beeinträchtigter Gesichtsnervenfunktion und (3) die Klassifizierung von HB 1-3. Insgesamt konnte eine gute Differenzierungsfähigkeit des mit Machine Learning geschätzten HB im Vergleich zum klinisch ermittelten HB gezeigt werden sowie die Fähigkeit klinisch relevante Unterschiede in der Fazialisfunktion zu detektieren. Somit hat das sEMG das Potential dazu, als objektivere Alternative zum HBGS zu dienen. In der zweiten retrospektiven Studie wurden multimodale Daten von 200 Patienten analysiert, u. a. intraoperative EMG-Aufzeichnungen, Traintime und A-Train-Cluster. Neuronale Feed-Forward-Netzwerke wurden mit diesen multimodalen Datensätzen trainiert, um die postoperative Fazialisfunktion (den klinischen HB-Graden entsprechend) auf der Grundlage dieser Faktoren zu schätzen. Die neuronale Netzwerkanalyse zeigte eine zufriedenstellende Leistung bei der Einschätzung der postoperativen HB-Grade. Insbesondere Netzwerke, die auf den Faktoren Traintime, präoperativer HB-Grad und Tumorgröße basierten, erzielten präzisere Resultate als solche, die nur anhand der Tumorgröße oder der mittleren Traintime die Fazialisfunktion einschätzten. Auf der Grundlage dieser Studiendaten können die Methoden des sEMG und der neuronalen Netzwerkanalyse in weiteren konfirmatorischen Studien eingesetzt werden, um die Erfassung der Fazialisfunktion in Echtzeit, insbesondere für den Einsatz in der klinischen Forschung, nachhaltig zu verbessern und hier eine objektive exakte Messmethodik zu ermöglichen.

Table of content

Abbreviations

ABR: Auditory Brainstem Response

CPA: Cerebellopontine angle

(s)EMG: (surface) Electromyography

FSRT: Fractionated Stereotactic Radiotherapy

HB: House Brackmann

HBGS: House Brackmann Grading System

MRI: Magnetic Resonance Imaging

QoL: Quality of Life

RCT: Randomized Controlled Trial

SRS: Stereotactic Radiosurgery

VS: Vestibular Schwannoma

1 Introduction

1.1 Vestibular schwannoma

Vestibular schwannomas (VS), also known as acoustic neurinomas, are brain tumors located in the cerebellopontine angle (CPA). As benign neoplasms they grow from Schwann cells arising from the vestibular component of the vestibulocochlear nerve (VIII). Despite a very small risk of malignant transformation, their displacing growth can endanger the function of intracranial structures and lead to clinical deficits[1]. Growth occurs first in the internal auditory canal which can lead to compression of the vestibular and cochlear nerve. The tumor progressively expands into the CPA and can affect the facial and trigeminal nerve here as well. The internationally established Koos grading system classifies VS tumor size based on extrameatal extension and brainstem compression into four grades[2]. This classification, presented below, was demonstrated to be a robust method for vestibular schwannoma categorization in a recently published reliability study by Erickson et al.[3].

Table 1: Koos grading system[2]

Koos Grade	**Description**
I	Intracanalicular
II	Extension into cerebellopontine angle, < 2 cm
III	Occupies cerebellopontine angle, no brainstem displacement, < 3 cm
IV	Brainstem displacement, > 3 cm

VS account for about 8% of all intracranial tumors and comprise over 80% of tumors in the CPA[4]. The overall incidence is 1-2/100.000/year, rising in recent years[5, 6]. In more than 95% of cases, VS occur unilaterally, in hereditary neurofibromatosis type 2, bilateral occurrence is possible[7]. Growth of the tumor is slow, and symptoms therefore often appear after years when the tumor mass begins to exert mechanical pressure on relevant intracranial structures. Typical symptoms are progressive hearing loss (90%), tinnitus (>60%), vertigo and, over time, facial nerve palsy (6%)[8].

Diagnosis is most often initiated based on otological and neurological symptoms. Otological symptoms are more common based on the beginning tumor growth in the inner auditory canal. The gold standard investigation tool for patients with unilateral hearing loss and tinnitus is magnetic resonance imaging (MRI). In a systematic review published in 2009 by Fortnum et al. the role of MRI as an initial screening tool in the identification of suspected VS was investigated regarding cost-effectiveness ratio. In comparison with auditory brainstem response (ABR)

sensitivity of T2-weighted imaging was high and relatively homogenous. Based on a cost-effectiveness model, the non-contrast MRI T2-weighted and T2-star-weighted showed to be the best and less costly option of first imaging screening in patients with suspicion of VS for accurate evaluation of the VIIIth and VIIth cranial nerves within the cerebellopontine angle and internal auditory canal followed by GdT1W if the suspected diagnosis is confirmed[9]. Recently, the European Association of Neuro-Oncology (EANO) recommended contrast-enhanced T1-weighted imaging as the method of choice for the final diagnosis and surgical planning. Commonly, VS appear isointense in the T1-weighted sequence with significant enhancement after gadolinium application. It presents as a solid mass, sometimes with cystic components, extending into the inner auditory canal which leads to its widening. Computer tomography and CISS MRI sequences can be used preoperatively to assess surgical anatomy of the skull base[10].

Various conservative and invasive treatment options exist for VS. Here, the tumor growth, clinical symptoms as well as the burden of the patient's disease must be weighed critically with regard to a treatment decision. Due to the slow growth rate a “watch and wait” strategy is feasible in case of asymptomatic patients which includes regular MRI scans and monitoring of hearing function via audiometry[11]. Current evidence shows no clear cut-off when to switch from a conservative strategy to invasive therapeutic interventions. However, observational studies with limited sample sizes identified extracanalicular components and diagnosis at young age as risk factors for an increased progression[12, 13]. Stereotactic radiosurgery with gamma knife technology (SRS) or fractionated stereotactic radiotherapy (FSRT) are the commonly used forms of radiotherapy treatment. SRS and FSRT are mainly used for smaller tumors (<3cm) in order to stop growth[14-17]. There exist no RCTs comparing SRS and FSRT. Six non-randomized trials however showed similar hearing function regarding both treatment options but a higher rate of facial palsy and trigeminal nerve dysfunction after FSRT[18-23].

In surgical therapy approaches, cure and therefore complete tumor resection presents the primary goal since residual tumor cells correlate with rate of recurrence. The decision for surgery should take tumor size, morphology, clinical symptoms, comorbidities, and the patient’s preferences into account[24]. Especially in Koos Grade IV, surgical resection is the treatment of choice to remove the symptomatic tumor mass and to protect from potentially life-threatening complications due to mechanical pressure. Patients should be treated in high volume centers, associated with a low surgery-related mortality of approximately about 0.5%[25]. The suboccipital retrosigmoid surgical approach is commonly favored because complete tumor resection of various sizes and location is feasible as well as preservation of facial nerve and hearing function. The brainstem, cranial nerves and relevant vascular structures can be visualized and identified clearly. The patient can be operated in a semi-sitting or horizontal position, whereas some small retrospective studies show an advantage of the semi-sitting position regarding functional outcomes[26-28]. Other

surgical approaches, such as the translabyrinthine or the middle fossa approach are feasible as well. However, the translabyrinthine approach requires labyrinthectomy which results in total loss of function of the inner ear. In comparison, the middle fossa approach may be an option for patients with small tumor size with improved postoperative hearing outcome. On the other hand, it leads to increased facial nerve manipulation due to the course of the nerve in relation to the operative access route[29]. For preservation of nerval function, intraoperative monitoring is used, including somatosensory evoked potentials, monitoring of the facial nerve comprising direct electrical stimulation, free-running EMG and EMG of the lower cranial nerves especially in case of Koos III and IV. Tumor size >3cm was identified as a predictive factor for the occurrence of postoperative facial palsy and associated with a worse outcome of nerval function one year after surgery[30]. Overall recurrence rates after surgical resection are about 14.5% mainly depending on the presence of residual tumor volume[31].

To sum up, the current evidence level regarding all therapeutic approaches for VS is low and inconclusive especially for Koos II and III. For small asymptomatic tumors, conservative observation is justifiable. In case of tumor growth and progressive clinical symptoms, radiotherapeutic and surgical approaches show similar results in terms of functional outcomes[32]. However, the surgical approach is the only treatment option that offers complete removal and thus cure of the tumor.

1.2 Facial nerve function

The facial nerve is the VII. cranial nerve and additionally contains fibers from the intermediate nerve. Its motor part innervates the muscles for facial expression, partly the muscles of the floor of the mouth and the stapedius muscle. The intermediate nerve part is responsible for taste sensation of the front two-thirds of the tongue through neuronal projections form the intermedial nerve. Parasympathetic fibers also innervate the lacrimal gland as well as several salivary glands.

The facial nerve originates in the brainstem and enters the temporal bone via the internal acoustic meatus. During its course through the temporal bone, the intermediate nerve part branches off fibers for gland and tongue innervation before exiting through the stylomastoid foramen. After its exit, the facial nerve itself branches out with its extracranial part into the motor fibers for the innervation of the facial muscular system[33, 34]. However, it has also been shown that the intermediate nerve carries motor fibers for the perioral muscles in addition to its sensitive, sensory, and parasympathetic parts[35, 36].

Facial nerve impairment can be caused by different pathologies and potentially impacts facial expression, lacrimation, salivation, taste, and sensation. It occurs most often unilaterally and is

classified into central and peripheral paralysis. Central nerve injury, commonly caused by stroke affecting the motor cortex due to ischemic or hemorrhagic origin, manifests in contralateral palsy of the oro-facial motor and sensory functions. Eyebrow and forehead movement are still possible due to the efferent input from both cerebral hemispheres[37]. Tumors and inflammatory diseases of the brain, such as multiple sclerosis, can also be the origin of central facial nerve palsy. Because of the complexity of the VIIth cranial nerve, peripheral nerve palsy manifests in different symptom complexes depending on the location of the lesion. The idiopathic form (Bell's palsy) is the most common form and hypotheses of the pathophysiology include herpes simplex virus infection as well as autoimmune inflammatory reactions[38]. In addition to the inflammatory and autoimmune genesis, paresis can also be caused by trauma such as fractures of the petrous bone or skull base as well as tumors of the cerebellopontine angle. Beside the motor component, including forehead, eye and mouth movement, hearing, secretion of tears, salivation and taste are affected as well. The typical clinical appearance of a peripheral nerve palsy consists of lacking wrinkling of the forehead, low eyebrow position, ptosis, hanging corner of the mouth and flattened nasolabial fold.

In case of benign slow growing neoplasms as the VS, facial function impairment often arises over the course of increasing tumor growth with pressure on the facial nerve in the CPA.

The quality of life in patients with facial function impairment depends on the prognosis and probability of recovery on the one hand, and on the severity of the paresis on the other. Especially incomplete eye closure with consecutive eye dryness plays a role. Psychological difficulties as well as restrictions of facial expression have a major impact on the patients' interpersonal relationships. A 2018 published retrospective study by Tavares-Brito et al. confirmed the correlation between facial palsy severity and QoL in 920 patients with various origins: Several other factors as duration of palsy, malignant and congenital etiologies were identified to correlate with higher QoL, possibly due to adaption to the facial impairment over time. Viral etiology, overweight, anxiety, chronic pain, previous treatment, and radiotherapy were associated with lower QoL[39]. A meta-analysis, which investigated the association between clinician assessed facial function and patient reported QoL[40] showed that QoL can only be moderately estimated based on facial function. Nevertheless, physical function had a greater impact on QoL than social function.

In VS patients, reasons for reduced quality of life (QoL) are multifactorial. They mainly report about a lack of energy, anxiety, headache and balance problems as psychosocial symptoms of their disease[41]. Tumor size seems to be one predictive factor for decreased QoL, but its importance depends on the different therapeutic strategies[42-44]. Since various factors play a role, studies comparing the different therapy methods show incongruent results[15, 43]. All three

therapy options, (1) conservative observation, (2) radiation and (3) microsurgery, show a comparable long-term outcome in terms of QoL for smaller tumors (<3cm). Patients treated with (1) and (2) show stable findings throughout the whole follow-up period; after surgical interventions, as expected, there is an initial deterioration but significant improvement, especially regarding mental aspects, in the further long-term period. In the case of permanent facial paresis HB ≥ grade IV, QoL is significantly limited due to incomplete eyelid closure.

In this patient population, it can be concluded that not only the therapy strategy plays a major role for long term quality of life but also patient related factors and that the therapy decision must therefore be made based on this multifactorial framework.

1.3 Facial nerve grading systems

Facial nerve grading systems serve to classify facial paresis for scientific and clinical assessment. Classification of the facial function is indispensable to monitor the clinical course as well as the effect of therapeutic interventions. Currently, the House-Brackmann facial nerve grading system (HBGS) is the standard method to assess the motor facial nerve function[45]. First introduced in 1983, it was endorsed as the main grading system by the Facial Nerve Disorders Committee of the American Academy of Otolaryngology-Head and Neck Surgery in 1984. The HBGS includes a 6-point scale with Grade I representing a normal facial function and Grade VI corresponding to total paralysis. The Grade is determined by clinical examination which includes the forehead (eyebrow movement), eye closure and movement of the corner of the mouth in comparison to the contralateral site. The assessment is done at rest and in motion. Initially, Brackmann developed a scale of 0.25cm divisions with a total score of 8 points (1-4cm for the mouth and eyebrow each) but over time, clinical practice favored the 6 grades system. Of utmost clinical relevance is the difference between HB grade III and IV, where eye closure is not completely possible anymore. As a result, irritations and injuries of the cornea occur more frequently and can even lead to totally unilateral blindness[46].

Table 2: the HBGS grading scale[47]

HBGS Grade	Description	Characteristics at rest	Characteristics in motion
I	Normal	Normal	Normal
II	Mild dysfunction	Normal symmetry & tone	Forehead: moderate/ good function; eye: complete closure with minimum effort; mouth: slight asymmetry
III	Moderate dysfunction	Normal symmetry & tone	Forehead: slight to moderate function; eye: complete closure with

			effort; mouth: slightly weak with maximum effort
IV	Moderately severe dysfunction	Normal symmetry & tone	Forehead: none; eye: incomplete closure; mouth: asymmetric with maximum effort
V	Severe dysfunction	Asymmetry	Forehead: none; eye: incomplete closure; mouth: slight movement
VI	Total paralysis	Asymmetry	Overall no movement

However, since its introduction, the aspects of reliability and suitability have been discussed constantly[48, 49]. Subjectivity due to clinical examination by only one assessor as well as accompanied interrater variability are named as the main points of criticism. Next to HBGS, many other facial grading systems have been developed to evaluate facial nerve function more properly. The Yanagihara grading system for instance was already developed and applied in 1976 as a regional classification system measuring 10 separate aspects of facial function summarized into one total score[50]. New systems like the Nottingham[51], the Sydney[52], the Sunnybrook Facial Grading Systems[53] or the MoReSS[54] as a modification of the HBGS were evolved in the following years. The aims of these systems were also to minimize the two factors mentioned above, nevertheless all approaches obtain their grading by clinical evaluation, and therefore, a complete elimination of subjectivity is not possible. This remaining issue can also be seen in novel approaches utilizing more modern techniques like video-analysis, e.g., as developed by Banks et al.[55, 56]. Their system is based on video assessment, albeit using a detailed and robust approach, considering static, movement-related and synkinesis parameters. Another software-based scoring system was investigated in a development and feasibility study by Taeger et al.[57]. Their approach was to develop an app-based assessment tool with smartphone depth mapping cameras to easily assess facial nerve function in everyday clinical practice. Data of the iPhone's TrueDepth camera system was processed for an app and an algorithm creating a grading scale for unilateral facial paresis. Adapting the Stennert index[58], individual facial regions were matched regarding a defined weighting including functional and cosmetic aspects. The assessment takes place while performing three facial movements and an infrared photograph measures the depth data of the face. Accordingly, a facial index from 0-100% was calculated. In the recently confirmed feasibility study, their approach could show reproducible results and no significant intertest variability.

However, availability and simple operability of these systems remain problematic either because of time-consuming assessment or quite expensive and limited availability of soft- and hardware which prevents their wider adoption as a standard facial grading system for clinical routine.

1.4 Perioperative monitoring of facial nerve function

The assessment and monitoring of the facial nerve in the perioperative setting in VS-surgery is an essential tool to determine the nerve function during the clinical course of the disease. In this way, correlations can be made between changes of nerve function and, for example, specific intraoperative events. First and foremost, intraoperative monitoring methods aim to prevent potential nerval damage. Preoperatively, the facial nerve function is routinely determined by clinical examination via the HBGS. Surface EMG (sEMG) measurement is also possible for detecting potential subtle nerve damage. sEMG measures the activity of the muscle fibers which is generated by the electric potential of the muscle cells from the surface above the muscle on the skin[59]. The sEMG in our exploratory trial was performed pre- and postoperatively and EMG amplitudes from seven different facial poses were recorded, which correspond to the movements used in the clinical examination of the HBGS.

Intraoperatively, the facial nerve can be monitored using various techniques. There are several aims of intraoperative monitoring: (1) identification of the nerve and mapping of its course in the operation field, (2) indication of potentially damaging surgical maneuvers and (3) intraoperative prognosis of the postoperative functional outcome. Direct electrical nerve stimulation is one of the oldest methods of monitoring the facial nerve. The facial nerve is electrically activated by direct stimulation with an electrified surgical instrument[60, 61]. With the help of different current strengths and stimulation techniques, the nerve can be identified accurately, even if it has already been thinned out by the tumor. However, due to the fact that this is a method of discontinuous stimulation, the ability to warn of possibly dangerous events as well as the predictive power is limited.

In contrast to direct stimulation, facial motor evoked potentials (FMEP) offer the possibility of monitoring the facial nerve already before it is visualized[62]. FMEP correspond to electric signals from the descending motor pathways and are recorded transcranially. The motor response is recorded with needle electrodes from facial muscles. In addition, one ipsilateral extremity muscle should be measured as control response. False positive findings are often obtained due to the short distance between the stimulation and recording area. This is because facial muscles may partly be activated directly by the electrical stimulus. In addition, high stimulation intensities may lead to pronounced patient movement, which make microsurgical preparation more difficult. As the main advantage, this method however allows for an estimation of functional integrity of the whole facial motor system right from the beginning of surgery; direct visual identification of the facial nerve is not necessary. Moreover, the prediction applies to both early and long-term outcome of facial function[63-65].

Free-running EMG differs from the first two methods. It does not involve electrically triggered electrical phenomena but consists of a continuous needle EMG of the muscles innervated by the facial nerve. In the case of patients under anesthesia, a zero-line EMG is recorded interspersed with various activity patterns during surgical manipulation ("surgically evoked potentials").

Highly relevant in this context are patterns called "A-trains", recorded by means of the free running EMG. These high frequency, low amplitude patterns are directly related to damaging manipulation of the N. facialis and correlate with the incidence of postoperative facial palsy[66]. Since visual quantification of the potentially numerous A-train events is difficult, an automated analysis was developed to detect A-train patterns and provide the summed total quantity – the "traintime". This quantitative parameter can be captured in real time[67]. This method is limited by the occurrence of false positive cases, partially caused by the intermediate nerve. In patients where there is evidence of a an intermediate nerve taking a course separate from the facial nerve in the cerebellopontine angle – also called a "split nerve" - there may be increased traintime, which does not correlate with an impaired postoperative facial nerve function[68], potentially due to damage to this split nerve which does not seem to be functionally significant. These intermediate nerve A-train patterns do not differ in morphology and frequency from the facial nerve A-trains which makes direct differentiation much more difficult and has a limiting effect on the prognostic significance of the A-trains[69].

1.5 Machine Learning and neural networks

Machine learning is considered a part of artificial intelligence. Compared to classical statistics, the computer is not given a dataset and a fixed analysis algorithm from which it generates the output dataset, but rather the algorithm adapts with "experience". Several methods exist how an algorithm can be trained to independently recognize patterns and correlations for analytic purposes. Based on the input of large amounts of data, the system can improve without being explicitly programmed. Supervised learning and unsupervised learning are mentioned here as two major machine learning paradigms. In a supervised machine learning setting an exemplary data set is provided with the corresponding output value. The algorithm learns to associate input and outputs and thus generate its own estimate of the relationship between the input dataset and the output data. Consequently, new datasets not seen during training sessions can be fed into the system, generating output based on the recognized patterns. In contrast to this, in the unsupervised machine learning setting, data without a defined or known target variable is used; the training data is not labeled. The machine learning algorithm then attempts to structure or cluster the data, potentially leading to generalized patterns. In addition, there is the semi-supervised learning approach, which includes both aspects of the above-mentioned paradigms.

Neural networks also represent a branch of artificial intelligence and a family of machine learning algorithms. They have been developed to recognize patterns in multimodal data sets. Neuronal networks are organized in different layers. Between an input and output layer one or multiple hidden layers may be placed to perform the computations. These algorithms approximate neuronal networks of the human brain at least in principle, albeit in a coarse manner. They can integrate different data types and capture complex interactions. Neuronal networks with more than a single layer are termed “deep neural networks”. Their use is often referred to as “deep learning”. According to their organization and consequently the flow of information, neural networks can be classified into different types, for example feedforward or recurrent networks. In feedforward networks, signals always run “forward” from the input to the output layer. The information flow is pointed in only one direction without feedback mechanism. In contrast, recurrent networks contain synapses back to “earlier” layers closer to the input layer, mediating feedback information[70].

1.6 Objective

The overall goal of the presented work is to improve and objectify facial nerve monitoring in the perioperative VS-surgery setting. By determining facial nerve function with surface EMG as an objective and reproducible method, further improvement for examining the exact nerve function at different perioperative time points is to be achieved. Using neural networks with the possibility of considering various pre- and intraoperative parameters, improved determination of postoperative nerve function in real-time during surgery shall be made. Here, the specific influence of the presence of an intermediate nerve regarding postoperative nerve function, and its prognostic relevance in relation to the overall traintime will be investigated with this multimodal analysis tool.

1.7 Contribution of the doctoral student

The effort of the doctoral student M.H. in this clinical-scientific work consisted primarily in developing, conducting, and analyzing the exploratory sEMG study. The doctoral student was involved in the concept development of this study with the supervisors J.P. and S.R. and contributed to the development of the sEMG measurement method. In the implementation phase, all 59 measurements were carried out independently by the doctoral student. During the analysis of the data, the doctoral student performed the preprocessing. The methodology of the multimodal machine learning approaches was discussed with the supervisor S.R., who carried out the final machine learning analyses. The writing and submission of the manuscript was done by the

doctoral student herself with the support of the supervisors S.R. and J.P. Regarding the second study, the doctoral student was involved in the preprocessing of data, supported the analysis process and assisted in the development of the final manuscript which provides an essential part for the future development of facialis monitoring with neural networks with the methodology of the sEMG from the exploratory study.

2 Discussion

The two sub-projects described here originated from the same overarching DFG-funded project and publications were written at roughly the same time. Both papers deal with a separate aspect of the overall topic "facial nerve monitoring". In the first part of this section, the application of surface EMG for the grading of facial nerve function with use of machine learning techniques is discussed. In the second part, the use of neural networks in an intraoperative setting is outlined, particularly with regard to A-trains and the influence of the intermediate nerve. Finally, the perspective part discusses the interrelations between both works and how the overall results are planned to be used for future research projects.

2.1 sEMG for grading of facial nerve function

It has already been shown[71, 72] that the use of sEMG is suitable for capturing facial function, and in our work, we were able to show that this measurement method is also suitable for the classification of facial nerve function and the degree of severity of paresis in VS-surgery patients analogous to HBGS. The results of the different machine learning algorithms presented good overall differentiation capability especially for clinically relevant differences (normal vs. impaired facial nerve function (HB 1 vs. HB 2-6) = AUC of 0.76 (Logistic Regression), 0.68 (SVM) and 0.73 (KNN), differentiation of normal or slight vs. moderate impaired facial nerve function (HB 1 and 2 vs. HB 3-6) = AUC of 0.88 (Logistic Regression), 0.97 (SVM) and 0.89 (KNN), classification of HB grades 1-3 = AUC of 0.74 (Logistic Regression), 0.70 (SVM) and 0.78 (KNN)).

A significant advantage of recording facial function by using sEMG is the objectivity of the method itself. Although the HBGS system is easy to use and can detect clinically relevant differences in facial paresis, the aspects of interrater variability and subjectivity by the individual examiner are disadvantages/ aspects that cannot be eliminated. The HBGS was invented in 1985 in a way different from how it is applied nowadays. Initially, it included metric measurements that made up the total score. The superior movement of the mid-portion of the top of the eyebrow, and the lateral movement of the angle of the mouth from both face halves should be measured and scaled in cm. Each 0.25cm scores 1 point up to 1cm of each face halves with a maximum score of 8. Thus, the original measurement methodology and scaling is supposed to remedy subjective interrater variability. In current practice, even though the HBGS classification was modified in 2009[73] with additional regional scoring aspects, it is used according to the method shown in figure 2 which offers many possibilities for individual subjective interpretation. The resulting “interobserver variability” was investigated by Scheller et al. in 2017 in the context of a randomized multi-center phase III trial assessing the efficacy of prophylactic nimodipine

treatment in VS surgery[49, 74]. In the early postoperative course, in only 36% of the cases, the different examiners agreed on the same HB grade. In 45%, a grade difference of 1 HB grade was observed, in 17% a two-grade difference and in 2% a three grade difference was documented. Especially in severe facial paresis an increased variability was shown between the different observers. This heterogeneity provides additional impetus for improvement in assessment of the facial nerve function in a more objective way. The results of our explorative trial give first indication that sEMG can conceivably be an alternative objective measurement method here. For clinical neurophysiological and neurosurgical research, the exact and objective determination of the facial function is indispensable. In particular, accurate prediction of the postoperative facial function with multimodal pre- and intraoperative parameters in real-time, was our main goal of the overarching DFG project.

With respect to the overall goal of generating more objectivity, sEMG measurement approach and the analysis of its results need to be considered separately. As an automated method and when trained with enough representative data sets, the latter yields reproducible results independent from the examiner.

Nevertheless, our study also has some limitations which must be mentioned here. First, the comparatively small sample size confines the validity. Additionally, not all HB grades could be captured which can be attributed to the sample size and the patient population. VS patients rarely present with a severe facial palsy preoperatively and also postoperatively these are rare with the generally occurrence in only one fifth of the patients[75]. The presence of a severe facial palsy correlates moreover with the tumor size. In our population Koos 3 was the mean grade as well and according to this, few patients showed a critical paresis preoperatively.

To conclude, our pilot study with a comparatively small sample size laid the foundation for further exploratory and confirmatory trials. Feasibility and effectiveness of the approach could be demonstrated. It is planned to evaluate the performance of the presented sEMG method in a more detailed way with a larger and thus more representative number of patients.

2.2 Neural networks for estimation of facial palsy

The outcome of facial nerve function after vestibular schwannoma surgery is multi-factorial. Preoperative known factors such as tumor size, previous radiation and incipient functional nerval impairment play an important role here. Intraoperative nerve monitoring using various methods is indispensable for avoiding damage and predicting postoperative nerve function. The use of free-running EMG, DNS and MEP has a significant impact on the preservation of the facial nerve and thus on quality of life after tumor resection[63, 76].

The role of A-trains, which are recorded with the free-running EMG, has already been subject of various investigations[67, 77, 78]. The total quantity of A-trains, the "traintime", correlates with the occurrence and severity of postoperative facial palsy. However, it has been demonstrated that a significant proportion of false positive A-trains can occur which do not correlate with clinical impairment of facial function under certain conditions. This fact limits the significance of the A-trains in their predictive power. In the search for the cause and origin of these false positive cases, a correlation with the occurrence of an intermediate nerve (split nerve) was shown[36, 79]. Due to the fact that "intermediate nerve A-trains" and "facial nerve A-trains" show uniform appearance in the free-running EMG, the identification of relevant A-trains for the postoperative outcome becomes more difficult. To describe the influence of the intermediate nerve more precisely as a confounder and in order to potentially eliminate it, the correlation of the occurrence of A-train clusters (A-trains over a majority of EMG-channels within a short time frame) with the presence of a split nerve has already been investigated[69]. The multi-factorial determination of postoperative facial nerve outcome including these confounders by using neural networks was therefore the subject of our second study.

It could be shown that neural networks based on the parameters traintime, preoperative HB-grade and tumor size attained the best results regarding estimation of the postoperative HB-grade (chi^2 54.8). Using only a single feature such as tumor size or the mean traintime, neural network performance decreased significant (chi^2 30.6 and 31.9). Removing A-train-clusters from the overall detected traintime, in addition to the above-mentioned parameters, further improved the prediction capability. This seems to confirm the assumption that A-trains contributed by A-train clusters and thus, mainly caused by intermediate nerve fibers, have no significant influence on the postoperative outcome. By removing this confounding activity, the outcome estimation by neural networks can be improved. In contrast, the addition of information on whether an intermediate nerve or A-train clusters were present in general did not improve the predictive power.

In conclusion, these findings underline the hypothesis that A-train clusters have a comparatively lower clinical impact. However, the clear influence of a separate intermediate nerve on this could not be proven. Consequently, in future investigations, the impact of the intermediate nerve on intraoperative nerve monitoring and false positive A-trains need to be identified. However, the conclusion can already be drawn that removal of A-train cluster traintime is to be preferred in patients without a detected separate intermediate nerve.

Neural networks are able to integrate data generated at different time points (pre- and intraoperatively) to adequately estimate postoperative HB grades. Furthermore, they are also able to integrate information from different sources and techniques, e.g. FMEP. Correspondingly, this

study of 200 patients may support the development and improvement of multimodal monitoring approaches.

2.3 Perspectives

The perioperative assessment of facial nerve function in patients with VS surgery is a highly relevant research topic and continuous improvements have been made in recent years. The overarching goal of facial nerve monitoring is not only the prevention of surgical damage, but also accurate prediction of its postoperative function intraoperatively in real-time. Using monitoring with such capabilities, the surgical procedure can be adapted immediately and damage to the nerve can be avoided. This DFG project (PR1275/1-1) was dedicated to this goal and aimed for the identifyication and minimization of possible confounders. The study "Semi-automated Grading of Facial Nerve Function" examined the sEMG method as a possible measurement instrument for extra-operative objective detection of facial nerve function in order to provide an alternative approach to the currently established subjective HBGS method. Data assessed with the HBGS is thus far included in the neural network analysis, as this is how preoperative nerve function is routinely assessed to date. If the sEMG method can be verified by further studies, it could be included as an aspect in the multimodal neural network approach and thus replace the subjective assessment of facial nerve function.

One limitation that affects both studies is the question of the true gold standard against which to compare new measurement and analytical approaches. In the first investigation, the question was whether facial function can be categorized in a more precise way than the previous system. The comparison with the clinical assessment thus limits the significance in terms of objectivity. In the neural network analysis for estimation of postoperative facial palsy, results were compared to clinical HB evaluation which resulted in deviations by one and up to two HB degrees. This can be caused in part by the fact that the HB grading system contains subjectivity in its nature as well as individual variability via its practical application. Both approaches may contribute to an improvement of intraoperative facial monitoring in the future. In terms of applicability, both methods have been developed primarily for clinical research purposes because of the time and material resources required, which are not ubiquitously available.

In future research projects, the aim is to further investigate the impact of a separate intermediate nerve on postoperative clinical outcome. In addition, the correlation between the regional distribution of paresis and the spatial distribution of A-train activity should be investigated, which has not been done so far. This detailed regional distribution cannot be analyzed with the current HBGS system. Using the sEMG, which records muscle activity in facial regions individually,

these regional differences can be assessed and potentially be interpreted with respect to recorded A-train clusters. This may provide further insight into the significance of intermediate nerve A-trains.

3 References

1. Cushing, H., *Tumors of the Nervus Acusticus and the Syndrome of the Cerebellopontine Angle*. W.B. Saunders. 1917.
2. Koos, W.T., et al., *Neurotopographic considerations in the microsurgical treatment of small acoustic neurinomas.* J Neurosurg, 1998. **88**(3): p. 506-12.
3. Erickson, N.J., et al., *Koos Classification of Vestibular Schwannomas: A Reliability Study.* Neurosurgery, 2019. **85**(3): p. 409-414.
4. Babu, R., et al., *Vestibular schwannomas in the modern era: epidemiology, treatment trends, and disparities in management.* J Neurosurg, 2013. **119**(1): p. 121-30.
5. Withrow, D.R., et al., *Nonmalignant meningioma and vestibular schwannoma incidence trends in the United States, 2004-2017.* Cancer, 2021. **127**(19): p. 3579-3590.
6. Samii, M., et al., *Surgery of cerebellopontine lesions.* Berlin Heidelberg: Springer, 2013.
7. Gupta, V.K., A. Thakker, and K.K. Gupta, *Vestibular Schwannoma: What We Know and Where We are Heading.* Head Neck Pathol, 2020. **14**(4): p. 1058-1066.
8. Kaul, V. and M.K. Cosetti, *Management of Vestibular Schwannoma (Including NF2): Facial Nerve Considerations.* Otolaryngol Clin North Am, 2018. **51**(6): p. 1193-1212.
9. Fortnum, H., et al., *The role of magnetic resonance imaging in the identification of suspected acoustic neuroma: a systematic review of clinical and cost effectiveness and natural history.* Health Technol Assess, 2009. **13**(18): p. iii-iv, ix-xi, 1-154.
10. Goldbrunner, R., et al., *EANO guideline on the diagnosis and treatment of vestibular schwannoma.* Neuro Oncol, 2020. **22**(1): p. 31-45.
11. Germano, I.M., et al., *Congress of Neurological Surgeons Systematic Review and Evidence-Based Guidelines on the Role of Radiosurgery and Radiation Therapy in the Management of Patients With Vestibular Schwannomas.* Neurosurgery, 2018. **82**(2): p. E49-E51.
12. Halliday, J., et al., *An update on the diagnosis and treatment of vestibular schwannoma.* Expert Rev Neurother, 2018. **18**(1): p. 29-39.
13. Ogawa, K., et al., *The growth rate of acoustic neuromas.* Acta Otolaryngol Suppl, 1991. **487**: p. 157-63.
14. Myrseth, E., et al., *Vestibular schwannoma: surgery or gamma knife radiosurgery? A prospective, nonrandomized study.* Neurosurgery, 2009. **64**(4): p. 654-61; discussion 661-3.
15. Regis, J., et al., *Functional outcome after gamma knife surgery or microsurgery for vestibular schwannomas.* J Neurosurg, 2002. **97**(5): p. 1091-100.
16. Karpinos, M., et al., *Treatment of acoustic neuroma: stereotactic radiosurgery vs. microsurgery.* Int J Radiat Oncol Biol Phys, 2002. **54**(5): p. 1410-21.

17. Pollock, B.E., et al., *Outcome analysis of acoustic neuroma management: a comparison of microsurgery and stereotactic radiosurgery.* Neurosurgery, 1995. **36**(1): p. 215-24; discussion 224-9.
18. Anderson, B.M., et al., *Single institution experience treating 104 vestibular schwannomas with fractionated stereotactic radiation therapy or stereotactic radiosurgery.* J Neurooncol, 2014. **116**(1): p. 187-93.
19. Andrews, D.W., et al., *Stereotactic radiosurgery and fractionated stereotactic radiotherapy for the treatment of acoustic schwannomas: comparative observations of 125 patients treated at one institution.* Int J Radiat Oncol Biol Phys, 2001. **50**(5): p. 1265-78.
20. Collen, C., et al., *Single fraction versus fractionated linac-based stereotactic radiotherapy for vestibular schwannoma: a single-institution experience.* Int J Radiat Oncol Biol Phys, 2011. **81**(4): p. e503-9.
21. Combs, S.E., et al., *Long-term outcome after highly advanced single-dose or fractionated radiotherapy in patients with vestibular schwannomas - pooled results from 3 large German centers.* Radiother Oncol, 2015. **114**(3): p. 378-83.
22. Kopp, C., et al., *Stereotactic fractionated radiotherapy and LINAC radiosurgery in the treatment of vestibular schwannoma-report about both stereotactic methods from a single institution.* Int J Radiat Oncol Biol Phys, 2011. **80**(5): p. 1485-91.
23. Meijer, O.W., et al., *Single-fraction vs. fractionated linac-based stereotactic radiosurgery for vestibular schwannoma: a single-institution study.* Int J Radiat Oncol Biol Phys, 2003. **56**(5): p. 1390-6.
24. Rutherford, S.A. and A.T. King, *Vestibular schwannoma management: What is the 'best' option?* Br J Neurosurg, 2005. **19**(4): p. 309-16.
25. McClelland, S., 3rd, et al., *Operative Mortality Rates of Acoustic Neuroma Surgery: A National Cancer Database Analysis.* Otol Neurotol, 2017. **38**(5): p. 751-753.
26. Rath, G.P., et al., *Complications related to positioning in posterior fossa craniectomy.* J Clin Neurosci, 2007. **14**(6): p. 520-5.
27. Roessler, K., et al., *Improved Postoperative Facial Nerve and Hearing Function in Retrosigmoid Vestibular Schwannoma Surgery Significantly Associated with Semisitting Position.* World Neurosurg, 2016. **87**: p. 290-7.
28. Spektor, S., et al., *Comparison of outcomes following complex posterior fossa surgery performed in the sitting versus lateral position.* J Clin Neurosci, 2015. **22**(4): p. 705-12.
29. DeMonte, F. and P.W. Gidley, *Hearing preservation surgery for vestibular schwannoma: experience with the middle fossa approach.* Neurosurg Focus, 2012. **33**(3): p. E10.
30. Killeen, D.E., et al., *The Association of Vestibular Schwannoma Volume With Facial Nerve Outcomes After Surgical Resection.* Laryngoscope, 2021. **131**(4): p. E1328-E1334.

31. Abouzari, M., et al., *Prediction of vestibular schwannoma recurrence using artificial neural network.* Laryngoscope Investig Otolaryngol, 2020. **5**(2): p. 278-285.
32. Neves Cavada, M., et al., *Intracanalicular Vestibular Schwannoma: A Systematic Review and Meta-analysis of Therapeutics Outcomes.* Otol Neurotol, 2021. **42**(3): p. 351-362.
33. Kochhar, A., B. Larian, and B. Azizzadeh, *Facial Nerve and Parotid Gland Anatomy.* Otolaryngol Clin North Am, 2016. **49**(2): p. 273-84.
34. Yang, S.H., et al., *Microsurgical anatomy of the facial nerve.* Clin Anat, 2021. **34**(1): p. 90-102.
35. Ashram, Y.A., et al., *Intraoperative electrophysiologic identification of the nervus intermedius.* Otol Neurotol, 2005. **26**(2): p. 274-9.
36. Alfieri, A., et al., *The relationship between nervus intermedius anatomy, ultrastructure, electrophysiology, and clinical function. Usefulness in cerebellopontine microsurgery.* Acta Neurochir (Wien), 2014. **156**(2): p. 403-8.
37. Schimmel, M., et al., *Oro-facial impairment in stroke patients.* J Oral Rehabil, 2017. **44**(4): p. 313-326.
38. Heckmann, J.G., et al., *The Diagnosis and Treatment of Idiopathic Facial Paresis (Bell's Palsy).* Dtsch Arztebl Int, 2019. **116**(41): p. 692-702.
39. Tavares-Brito, J., et al., *Facial Palsy-Specific Quality of Life in 920 Patients: Correlation With Clinician-Graded Severity and Predicting Factors.* Laryngoscope, 2019. **129**(1): p. 100-104.
40. Bruins, T.E., et al., *Associations Between Clinician-Graded Facial Function and Patient-Reported Quality of Life in Adults With Peripheral Facial Palsy: A Systematic Review and Meta-analysis.* JAMA Otolaryngol Head Neck Surg, 2021. **147**(8): p. 717-728.
41. Pruijn, I.M.J., et al., *What determines quality of life in patients with vestibular schwannoma?* Clin Otolaryngol, 2021. **46**(2): p. 412-420.
42. Carlson, M.L., et al., *Long-term quality of life in patients with vestibular schwannoma: an international multicenter cross-sectional study comparing microsurgery, stereotactic radiosurgery, observation, and nontumor controls.* J Neurosurg, 2015. **122**(4): p. 833-42.
43. Di Maio, S. and R. Akagami, *Prospective comparison of quality of life before and after observation, radiation, or surgery for vestibular schwannomas.* J Neurosurg, 2009. **111**(4): p. 855-62.
44. Foley, R.W., et al., *The Impact of Primary Treatment Strategy on the Quality of Life in Patients with Vestibular Schwannoma.* World Neurosurg, 2017. **102**: p. 111-116.
45. House, J.W. and D.E. Brackmann, *Facial nerve grading system.* Otolaryngol Head Neck Surg, 1985. **93**(2): p. 146-7.
46. Rahman, I. and S.A. Sadiq, *Ophthalmic management of facial nerve palsy: a review.* Surv Ophthalmol, 2007. **52**(2): p. 121-44.

47. Sun, M.Z., et al., *Neuroanatomical correlation of the House-Brackmann grading system in the microsurgical treatment of vestibular schwannoma.* Neurosurg Focus, 2012. **33**(3): p. E7.
48. Kang, T.S., et al., *Facial nerve grading systems (1985-2002): beyond the House-Brackmann scale.* Otol Neurotol, 2002. **23**(5): p. 767-71.
49. Scheller, C., et al., *Interobserver variability of the House-Brackmann facial nerve grading system for the analysis of a randomized multi-center phase III trial.* Acta Neurochir (Wien), 2017. **159**(4): p. 733-738.
50. Y, N., *Facial Nerve Surgery, Proceedings: Third International Symposium on Facial Nerve Surgery.* Fisch U, ed, 1977: p. pp: 533–535.
51. Murty, G.E., et al., *The Nottingham System: objective assessment of facial nerve function in the clinic.* Otolaryngol Head Neck Surg, 1994. **110**(2): p. 156-61.
52. Coulson, S.E., et al., *Reliability of the "Sydney," "Sunnybrook," and "House Brackmann" facial grading systems to assess voluntary movement and synkinesis after facial nerve paralysis.* Otolaryngol Head Neck Surg, 2005. **132**(4): p. 543-9.
53. Ross, B.G., G. Fradet, and J.M. Nedzelski, *Development of a sensitive clinical facial grading system.* Otolaryngol Head Neck Surg, 1996. **114**(3): p. 380-6.
54. de Ru, J.A., et al., *Grading facial nerve function: why a new grading system, the MoReSS, should be proposed.* Otol Neurotol, 2006. **27**(7): p. 1030-6.
55. Banks, C.A., et al., *Clinician-Graded Electronic Facial Paralysis Assessment: The eFACE.* Plast Reconstr Surg, 2015. **136**(2): p. 223e-230e.
56. Banks, C.A., N. Jowett, and T.A. Hadlock, *Test-Retest Reliability and Agreement Between In-Person and Video Assessment of Facial Mimetic Function Using the eFACE Facial Grading System.* JAMA Facial Plast Surg, 2017. **19**(3): p. 206-211.
57. Taeger, J., et al., *Utilization of Smartphone Depth Mapping Cameras for App-Based Grading of Facial Movement Disorders: Development and Feasibility Study.* JMIR Mhealth Uhealth, 2021. **9**(1): p. e19346.
58. Stennert, E., C.H. Limberg, and K.P. Frentrup, *[An index for paresis and defective healing--an easily applied method for objectively determining therapeutic results in facial paresis (author's transl)].* HNO, 1977. **25**(7): p. 238-45.
59. Wozniak, K., et al., *Surface electromyography in orthodontics - a literature review.* Med Sci Monit, 2013. **19**: p. 416-23.
60. Delgado, T.E., et al., *Intraoperative monitoring of facila muscle evoked responses obtained by intracranial stimulation of the facila nerve: a more accurate technique for facila nerve dissection.* Neurosurgery, 1979. **4**(5): p. 418-21.
61. Kartush, J.M., et al., *Intraoperative facial nerve monitoring: a comparison of stimulating electrodes.* Laryngoscope, 1985. **95**(12): p. 1536-40.

62. Dong, C.C., et al., *Intraoperative facial motor evoked potential monitoring with transcranial electrical stimulation during skull base surgery.* Clin Neurophysiol, 2005. **116**(3): p. 588-96.
63. Prell, J., et al., *Intraoperative monitoring of the facial nerve : Vestibular schwannoma surgery.* HNO, 2017. **65**(5): p. 404-412.
64. Matthies, C., et al., *Facial motor evoked potentials in cerebellopontine angle surgery: technique, pitfalls and predictive value.* Clin Neurol Neurosurg, 2011. **113**(10): p. 872-9.
65. Acioly, M.A., et al., *Transcranial electrocortical stimulation to monitor the facial nerve motor function during cerebellopontine angle surgery.* Neurosurgery, 2010. **66**(6 Suppl Operative): p. 354-61; discussion 362.
66. Romstock, J., C. Strauss, and R. Fahlbusch, *Continuous electromyography monitoring of motor cranial nerves during cerebellopontine angle surgery.* J Neurosurg, 2000. **93**(4): p. 586-93.
67. Prell, J., et al., *A real-time monitoring system for the facial nerve.* Neurosurgery, 2010. **66**(6): p. 1064-73; discussion 1073.
68. Prell, J., et al., *The intermedius nerve as a confounding variable for monitoring of the free-running electromyogram.* Clin Neurophysiol, 2015. **126**(9): p. 1833-9.
69. Rampp, S., et al., *A-train clusters and the intermedius nerve in vestibular schwannoma patients.* Clin Neurophysiol, 2019. **130**(5): p. 722-726.
70. Kriegeskorte, N. and T. Golan, *Neural network models and deep learning.* Curr Biol, 2019. **29**(7): p. R231-R236.
71. Kim, J.U., et al., *A Study on the Correlation between Surface Electromyography and Assessment Scale for Facial Palsy.* The Acupuncture, 2013. **30**(5): p. 107-116.
72. Ryu, H.M., et al., *Study on the Validity of Surface Electromyography as Assessment Tools for Facial Nerve Palsy.* J Pharmacopuncture, 2018. **21**(4): p. 258-267.
73. Vrabec, J.T., et al., *Facial Nerve Grading System 2.0.* Otolaryngol Head Neck Surg, 2009. **140**(4): p. 445-50.
74. Scheller, C., et al., *Prophylactic nimodipine treatment for cochlear and facial nerve preservation after vestibular schwannoma surgery: a randomized multicenter Phase III trial.* J Neurosurg, 2016. **124**(3): p. 657-64.
75. Taha, I., et al., *Facial nerve function and hearing after microsurgical removal of sporadic vestibular schwannomas in a population-based cohort.* Acta Neurochir (Wien), 2020. **162**(1): p. 43-54.
76. Stankovic, P., et al., *Continuous intraoperative neuromonitoring (cIONM) in head and neck surgery-a review. German version.* HNO, 2020. **68**(11): p. 801-809.

77. Prell, J., et al., *Facial nerve palsy after vestibular schwannoma surgery: dynamic risk-stratification based on continuous EMG-monitoring.* Clin Neurophysiol, 2014. **125**(2): p. 415-21.

78. Rampp, S., et al., *A-trains for intraoperative monitoring in patients with recurrent vestibular schwannoma.* Acta Neurochir (Wien), 2013. **155**(12): p. 2273-9; discussion 2279.

79. Alfieri, A., et al., *The nervus intermedius as a variable landmark and critical structure in cerebellopontine angle surgery: an anatomical study and classification.* Acta Neurochir (Wien), 2012. **154**(7): p. 1263-8.

4 Theses

1. Vestibular schwannomas and their surgical removal effect the facial nerve in several ways which makes an accurate assessment of its function necessary.
2. sEMG can be used to grade facial nerve function and may be a potential alternative objective assessment tool to HBGS, primarily for scientific investigations.
3. With neural networks based on preoperative HB, traintime and tumor size, accurate estimations of postoperative HB can be achieved.
4. The removal of A-train clusters, associated with clinically not relevant A-trains generated by the intermediate nerve, improves the predictive power of estimation of postoperative facial nerve function.
5. Perioperative nerve monitoring methods including multimodal data sources and diagnostic methods may enable to capture the exact nerve function in real-time.

5 Publications

1 Holze M, Rensch L, Prell J, Scheller C, Simmermacher S, Scheer M, Strauss C, Rampp R. 2022 Oct. Learning from EMG: Semi-automated Grading of Facial Nerve Function. J Clin Monit Comput. 36(5):1509-1517.

2 Rampp S, Holze M, Scheller C, Strauss C, Prell J. 2022 Dec. Neural networks for estimation of facial palsy after vestibular schwannoma surgery. J Clin Monit Comput. Online ahead of print.

Journal of Clinical Monitoring and Computing (2022) 36:1509–1517
https://doi.org/10.1007/s10877-021-00793-y

ORIGINAL RESEARCH

Learning from EMG: semi-automated grading of facial nerve function

Magdalena Holze[1,2] · Leonhard Rensch[1] · Julian Prell[1] · Christian Scheller[1] · Sebastian Simmermacher[1] · Maximilian Scheer[1] · Christian Strauss[1] · Stefan Rampp[1]

Received: 25 June 2021 / Accepted: 19 December 2021 / Published online: 6 January 2022

Abstract
The current grading of facial nerve function is based on subjective impression with the established assessment scale of House and Brackmann (HB). Especially for research a more objective method is needed to lower the interobserver variability to a minimum. We developed a semi-automated grading system based on (facial) surface EMG-data measuring the facial nerve function of 28 patients with vestibular schwannoma surgery. The sEMG was recorded preoperatively, postoperatively and after 3–12 months. In addition, the HB grade was determined. After manual selection and preprocessing, the data were subjected to machine learning classificators (Logistic regression, SVM and KNN). Lateralization indices were calculated and multivariant machine learning analysis was performed according to three scenarios [differentiation of normal (1) and slight (2) vs. impaired facial nerve function and classification of HB 1-3 (3)]. The calculated AUC for each scenario showed overall good differentiation capability with a median AUC of 0.72 for scenario 1, 0.91 for scenario 2 and multiclass AUC of 0.74 for scenario 3. This study approach using sEMG and machine learning shows feasibility regarding facial nerve grading in perioperative VS-surgery setting. sEMG may be a viable alternative to House Brackmann regarding objective evaluation of facial function especially for research purposes.

Keywords House–Brackmann · Facial EMG · Facial nerve function · Grading system · Interobserver variability · Vestibular schwannoma

1 Introduction

Vestibular schwannomas are brain tumors located in the cerebellopontine angle (CPA) which affect the facial nerve in several ways. Its location may lead to compression of the facial nerve with potential impact on its function [1]. Furthermore, VS surgery itself may damage the facial nerve and results in postoperative facial palsy. This may have consequences for patient health and quality of life [2–4].

Currently, the House-Brackmann facial nerve grading system (HBGS) is the standard method to assess facial nerve function. First introduced in 1983 it was endorsed as the main grading system by the Facial Nerve Disorders Committee of the American Academy of Otolaryngology-Head and Neck Surgery in 1984 [5]. However, since then its reliability and suitability has been discussed constantly [6, 7].

Next to HBGS, many other facial grading systems have been developed to classify facial nerve function. The Yanagihara grading system for instance was already developed and applied in 1976 as a regional classification system measuring ten separate aspects of facial function summarized into a total score [8]. New systems like the Nottingham [9], the Sydney [10], the Sunnybrook Facial Grading Systems [11] or the MoReSS [12] as a modification of the HBGS were evolved in the following years.

Although all these grading systems and modifications have been developed, the common issue of observer subjectivity remains a significant source of variability and inaccuracy [6, 13]. Smith et al. compared multiple leading systems, including HBGS, and found similar interobserver variation in all of them [14]. The HBGS was updated in 2009 by the Facial Nerve Disorders Committee to incorporate a regional scoring scale and to limit interobserver variability [15].

✉ Magdalena Holze
magdalena.holze@med.uni-heidelberg.de

[1] Department of Neurosurgery, University Hospital Halle (Saale), Halle, Germany

[2] Present Address: Department for General, Visceral and Transplantation Surgery, University Hospital Heidelberg, Heidelberg, Germany

Nevertheless, this resulted in only moderate improvements and limited impact on clinical practice as Scheller et al. [6] showed in 2017. Despite its easy application and ability to identify clinically relevant facial palsy, the limited interrater reliability of also the updated HBGS impacts its sensitivity and robustness required to evaluate smaller differences.

These remaining issues can also be seen in ongoing publication of novel approaches utilizing more modern techniques like video-analysis, e.g. as developed by Banks et al. [16, 17]. Their system is based on subject video assessment, albeit using a detailed and robust approach, considering static, movement and synkinesis parameter. However, availability and simple operability of these systems remain problematic which prevents their wider adoption as a standard facial grading system for clinical routine.

The ability to document more subtle differences in facial nerve function with low interrater variability has limited relevance for clinical practice. However, this aspect becomes especially important for research. Large interrater variability significantly constrains development and optimization e.g. of neuromonitoring approaches and pharmacotherapy [18–20]. The surface electromyography (sEMG) based method in the current study is intended to meet such requirements of low examiner-dependence with limited added measuring effort.

Our approach evaluates facial nerve function using surface sEMG. sEMG measures the activation of muscle fibres which is related to the level of contraction of motor units. Ryu et al. and Kim et al. showed that there is indeed considerable correlation between sEMG and clinical assessment tools (House–Brackmann scale, Yanagihara grading system, Sunnybrook facial grading system) [21, 22]. Despite these studies on sEMG to assess facial function, there is a lack of studies on neurosurgical patients undergoing VS surgery.

Consequently, there is still a need for a new objective measurement technique for clinical research in this area with the aim of improving surgical treatment, specifically regarding facial nerve function. Our exploratory study shows that sEMG may be a potential solution.

2 Method

2.1 Patients

Twenty-eight patients undergoing elective surgery for vestibular schwannoma were recruited for the study from April 2017 until July 2018 in the order in which the patients had their surgery appointment or control investigation.

Inclusion criteria were (1) indication for vestibular schwannoma surgery due to suspected VS, independent of primary tumor or recurrence, tumor size and postoperative histological diagnosis and (2) adult age.

The exclusion criteria were (1) neurofibromatosis, (2) tumor at a different location with the consequence of other surgical procedures (e.g. ependymoma/ metastasis of the posterior cranial fossa with similar symptoms) and (3) preoperative facial palsy caused by central nerve disease.

The study was positively reviewed by the institutional review board of the University Hospital Halle (Saale). All patients gave their written informed consent to participate in the study.

2.2 Recordings

sEMG was recorded with a Grass-Telefactor 15LT biosignal amplifier (West Warwick, Rhode Island) which is also used for standard procedures, such as continuous EMG and intraoperative monitoring (IOM) at the Department of Neurosurgery at the University Hospital Halle (Saale) and already described in earlier studies [18].

sEMG was recorded at defined time points: 1 day before surgery, between the fifth and tenth postoperative day and at follow-up evaluations at 3, 6 or 12 months after surgery. EMG amplitudes from seven different facial poses were recorded (Fig. 1). These poses where chosen to ensure high comparability between measurements and clinical evaluation. They were the same movements as those used in the examination of the HBGS. In addition, they show the highest EMG activity of the representative facial nerve innervated muscles [23, 24].

One electrode pair was placed on the forehead, one on the nasolabial fold and one was positioned underneath the lips on the lateral chin (Fig. 1). A ground electrode was positioned on the right wrist. Before application of the electrodes, the skin was cleaned with alcohol swipes to reduce impedance.

Before the measurement, the movements were demonstrated by the examiner and the patients were encouraged to practice every movement to guarantee correct performance. The patients were also motivated to perform the movements as strongly as they could to record the maximum muscle activity. EMG activity was recorded during three repetitions for each pose to capture intraindividual variability. The tension time was about 1 s, the relaxation time 3 s. All measurements were done by the same person (author MH) who was fully familiar with the measurement method and instrument to prevent the effect of variability due to different examiners.

2.3 Assessment scale

The House–Brackmann Grade was determined by a single examiner (author JP) with the aid of photographs of all poses. These were routinely taken at all defined time points.

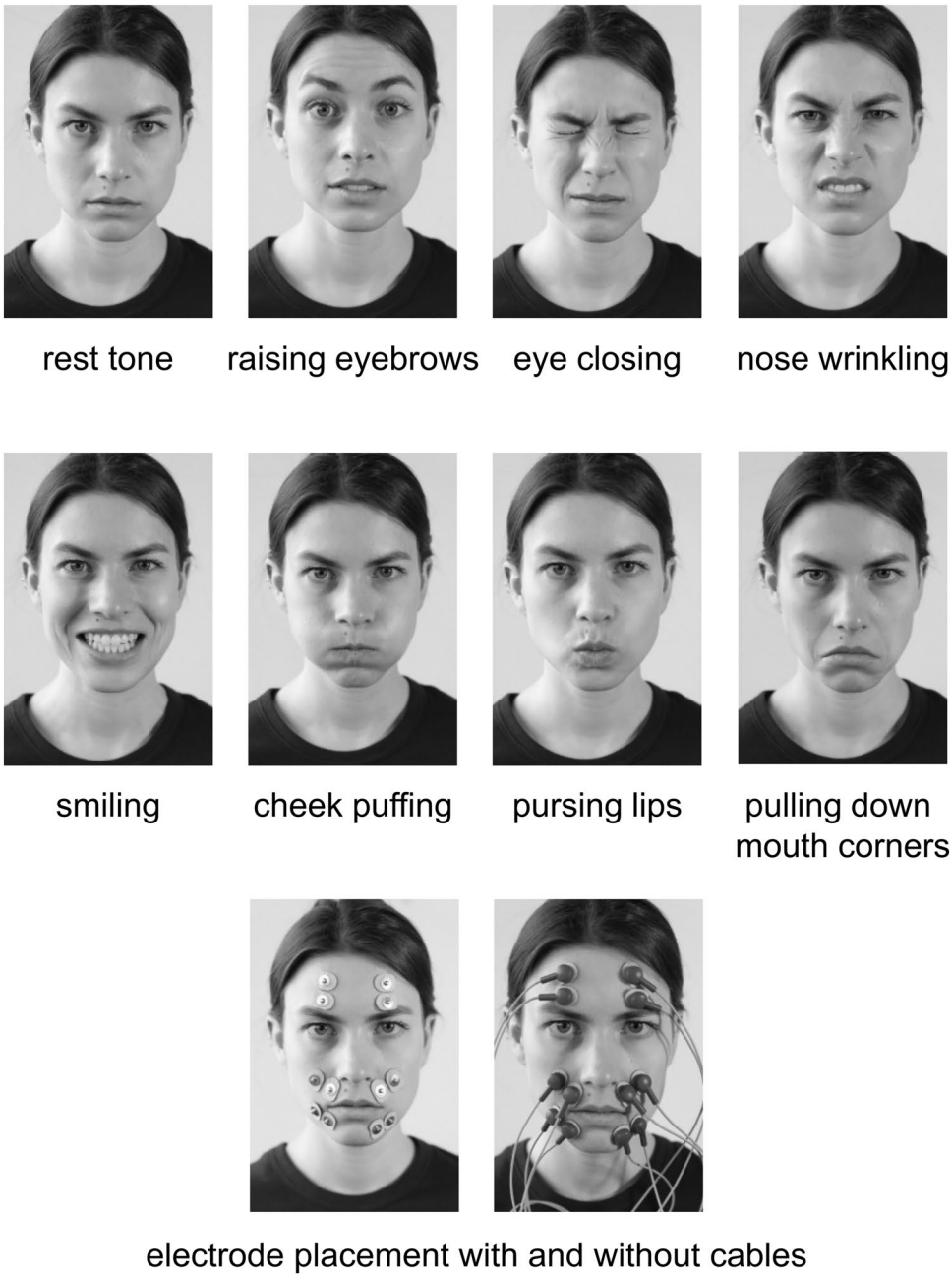

Fig. 1 Demonstration of the seven different facial poses and electrode placement

2.4 Data processing

As a first step, 500 ms epochs containing the maximum amplitudes of each movement were manually selected. (in-house software). Data were then rectified to obtain the absolute amplitude and smoothed with 100ms window running average. Calculation of the 95th percentile then yielded a single EMG amplitude value per channel. An overview of the various methodological and analytical steps in chronological order is shown in Fig. 2.

2.5 Statistical analysis

From the ipsi- and contralateral EMG amplitudes of the orbicular oculi, nasalis and orbicular oris muscles and each repetition, lateralization indices (LI) were calculated according to the formula:

$$LI = (EMG\_ipsi - EMG\_contra)/(EMG\_ipsi + EMG\_contra).$$

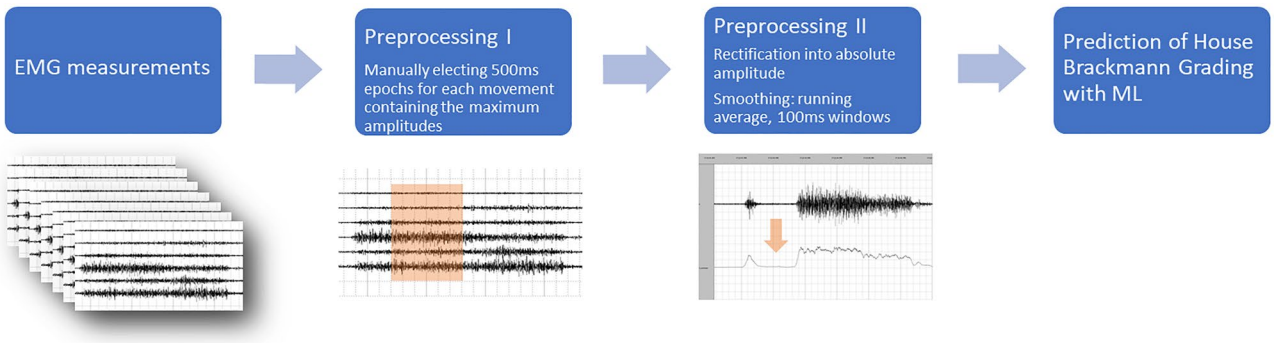

Fig. 2 Overview of the methodological and analytical steps

While absolute EMG amplitudes likely contain information about facial nerve function, intraindividual asymmetries of ipsi- vs. contralateral facial muscles strongly influence clinical evaluation of HB grades. Furthermore, LI take interindividual variability of facial muscle movement as well as small differences in recording setup into account. Due to the limited sample size we only used LI for analysis and did not additionally include absolute values. Furthermore, most HB grades were in the range of HB 1–3, which are characterized by subtler changes which might be more apparent by comparing ipsi- to contralateral EMG intraindividually (overview shown in Fig. 3).

Multivariate machine learning analysis was performed with scikit-learn version 0.24.2 (https://scikit-learn.org/), applying three algorithms: Logistic Regression, support vector machines (SVM using gamma = 1/number of features) and k nearest neighbor classification (KNN using k = 5 and distance weighting of neighbors). The aim was to take advantage of an automatic learning system which is able to recognize patterns between input and output data and to make predictions based on this process. To evaluate the viability of estimating HB grades from EMG data, we evaluated three classification scenarios: (1) Differentiation of normal vs. impaired facial nerve function (HB 1 vs. HB 2–6), (2) normal or slight vs. moderate impairment (HB 1–2 vs. HB 3–6) and (3) classification of HB grades 1–3. Further differentiation of individual classification of each possible HB grade (1–6) were not evaluated due to the limited sample size. Consequently, scenarios 1 and 2 represented a binary classification problem, whereas scenario 3 aims at multiclass classification. For the latter, the ordinal scaling of HB grades was not taken into account; multinomial classification was attempted. Due to the low sample size we did not optimize parameters of the three algorithms.

Cross-validation for training and testing utilized a "leave one out" strategy. This means that during each training-testing run, data from all patients except one were used for training, and testing was performed using the remaining. This was repeated until data from each patient had been left out and used for testing once, which aims to limit overfitting.

This training aims to let the respective algorithm learn the pattern between in- and output data, which can then be applied to estimate output data associated with never seen input data. The degree of correct estimated output data is then used to evaluate the performance. During the testing, the corresponding HB grades were estimated based on the learned patterns and compared to the actual HB grades.

Classification results, respectively class probabilities of only this testing split were accumulated and compared to the true clinical HB grades for estimation of performance. We calculated receiver-operator characteristic (ROC) and area-under-the-curve (AUC) values for scenarios 1 and 2. For scenario 3, the multiclass "one versus one" AUC was calculated, which computes the average AUC of all possible pairwise combinations of classes. Due to the largely imbalanced dataset with most patients with lower HB grades, we preferred AUC over accuracy due to the lower albeit still present susceptibility to this issue.

Additionally, an analysis was carried out to determine the over- and underestimation of each classifier compared to the clinical examined HB (for this purpose, only the test data from HB 1–3 of the leave-one-out strategy was used).

3 Results

3.1 Baseline data

Overall, 28 patients were recruited for sEMG-measurement. Mean age was 50 years (range 23–77); 75% were female. Mean tumor size was Koos 3 (range 1–6) [25] and 75% of the tumors were located on the left side. The postoperative histological examination revealed vestibular schwannoma in 24 patients (86%) and other histologies in four cases (14%, three meningiomas and one intermedius neurinoma). In a total of 59 measurements, 30 times a clinical HB 1, 17

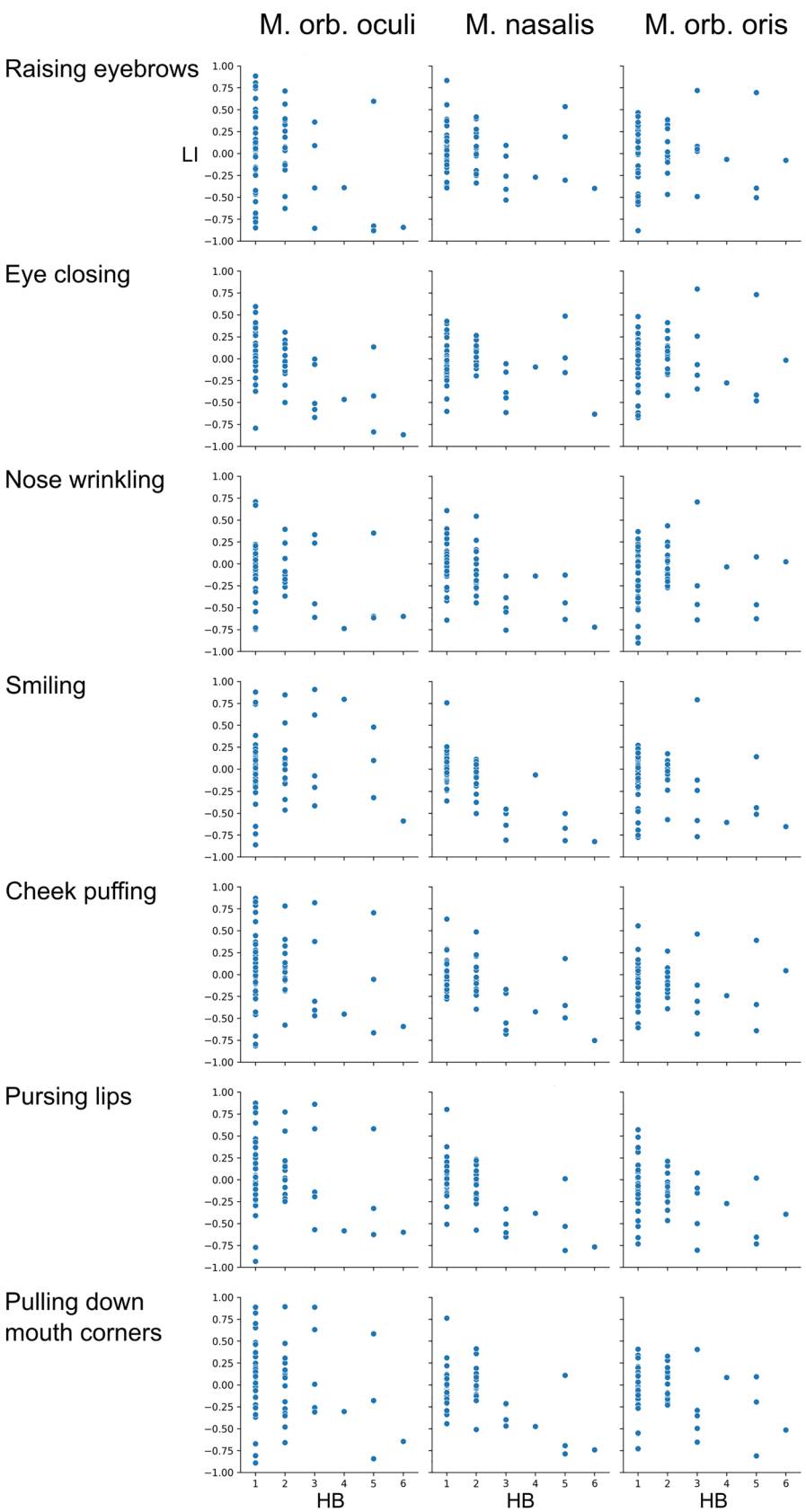

Fig. 3 Overview of the lateralization indices (LI) of the individual muscle groups for all sevenmovements for the respective HB grades. A LI of one would result from sEMG-activity only ipsilateral tothe operated side, respectively a LI of − 1 would show only contralateral sEMG

times a clinical HB 2, 5 times a HB 3, 3 times a HB 4 and 2 times a HB 5 was evaluated. There were no HB grade 6 measurements. In two measurements (1 preoperative and 1 postoperative, each from different patients) the corresponding clinically evaluated HB was missing.

3.2 Participant flow

There were no patient dropouts before preoperative data acquisition. In case of three patients, preoperative sEMG, in two patients the first postoperative sEMG and in 18 patients the long-term measurements were missing because of logistical reasons and limitation of the follow-up period. A total of 23 patients underwent both pre- and postoperative measurements, eight of them additionally during follow-up. Regarding House–Brackmann assessment, one preoperative evaluation was missing and one of the follow-ups.

In two cases of preoperative measurements, data of single repetitions and complete movements were missing because the patient could not perform the movement correctly or strong enough to achieve a representative EMG amplitude for further analysis. In one case smiling and blowing out cheeks was not possible, wrinkling the nose was accomplished just a single time. In the other case smiling was accomplished just one time instead of the three required trials. Such missing values were replaced with either the values of the available single measurement (nose wrinkling), so that there was virtually no mean value or the average of all other movements (in case of completely missing data of a movement). This procedure guarantees that the missing values are replaced by similar numbers instead of e.g. leaving those values out or by zeroing the missing values, which would distort the learning procedure.

3.3 Estimation of facial nerve function

The first and second scenarios showed the following AUC depending on each classificator: Evaluating the differentiation of normal vs. impaired facial nerve function (HB 1 vs. HB 2–6) resulted in an AUC of 0.76 (Logistic Regression), 0.68 (SVM) and 0.73 (KNN). Differentiation of normal or slight vs. moderate impaired facial nerve function (HB 1 and 2 vs. HB 3–6) showed an AUC of 0.88 (Logistic Regression), 0.97 (SVM) and 0.89 (KNN). The multiclass AUC of HB 1, 2, and 3 carried out for scenario 3 yielded in an AUC of 0.74 (Logistic Regression), 0.70 (SVM) and 0.78 (KNN) each.

In addition, under- and overestimation by each algorithm were analyzed with the test data of HB 1–3, scenario 3. Logistic Regression was concordant in 66% and underestimated in 20% of the cases, i.e. it predicted a "better" HB grade than the actual clinical HB. There were 14% overestimated cases by Logistic Regression. SVM showed a concordance of 63% and predicted a "better" HB than it was clinically evaluated in 36% of cases, and it predicted a "worse" HB in 2%. For KNN the concordance was 66%, underestimation 27% and the overestimation 7%.

4 Discussion

To the best of our knowledge this is the first study to use sEMG and machine learning as an objective assessment tool before and after VS-surgery for comparison with HBGS.

Our results show that there is an overall good differentiation capability of the estimated HB with machine learning compared to the clinical determined HB and clinically relevant differences can be distinguished.

4.1 Machine learning vs. clinical assessment

HB grades estimated by machine learning algorithms showed a good differentiability and especially clinically relevant gradings of facial nerve function could be distinguished. Differentiation between slight (HB 1–2) and moderate impaired facial nerve function (HB 3–6, scenario 2) showed with 0.88–0.97 the best results. In addition, the AUC of scenario one (0.68–0,73, differentiation of normal vs. impaired facial function) confirms an overall acceptable performance and motivates to further studies with larger data sets beyond the exploratory character of our current investigation.

A general concern when comparing the estimated results with the clinical assessment is the lack of a true gold standard. A main question of our new approach is whether it categorizes the facial function better and more precisely than the previous grading systems. While HB is the de-facto clinical standard, its evaluation is subjective and suffers from considerable interrater variability [6]. We tried to minimize interobserver variability as much as possible by having the same investigator determine the HB grade. A high or even perfect correlation between our method and HB is not expected nor desirable as this would require reproducing the subjective variability.

Logistic regression, SVM and KNN showed different degrees of under- and overestimation i.e. deviation from the clinically rated HB. For example, an examined HB 1 was partially rated as HB 2 (overestimation for Logistic regression of 14%, for SVM of 2% and for KNN of 7%) and vice versa, a HB 2 as HB 1 (Logistic regression underestimated in 20%, SVM in 36% and KNN in 27% of the cases). An underestimated HB could be caused by high EMG activity, despite already present palsy, caused by synkinesis and attempted compensation by the opposite side. An overestimated HB on the other hand could be due to subtle asymmetries in muscle activity which

is only detectable by EMG but not yet visible for the investigator. Alternatively, of course, fatigue or limited patient compliance are further explanations for overestimated HB.

In order to additionally illustrate the advantages of the machine learning approach regarding interrater reliability, we evaluated concordance rates with the clinical examiner. Logistic regression and KNN showed concordance with the clinical examiner in 66% respectively, SVM in 63%. After training, the classification results provided by all three approaches are guaranteed to be reproducible with the same input data, leaving factors of the sEMG recording, e.g. patient compliance, data quality, etc. as the only sources of variability.

In comparison, the study by Scheller et al. showed an overall concordance of only 36% and interobserver variability by one degree of 45%, by two degrees of 17% and by three degrees of 1% [6]. Reproducibility of clinical evaluation was not investigated. The study used patient photos for evaluation, i.e. the reported concordance rates already exclude variability from the clinical investigation itself. Overall, although our concordance data is restricted to the lower HB grades, our results suggest that a similar or improved performance with high reproducibility using sEMG should be viable.

4.2 Alternative methods

Many facial nerve grading systems have been developed in the past [5, 8–12], most of them with the goal to provide clinical assessment tools. All these systems share the approach that the clinical observer is the central evaluating component. The problem of subjectivity and interobserver variability is therefore inherent to such assessment tools and can only be improved to a certain extent [6, 7, 13, 14]. Their strength consists in the easy handling and practicability in clinical routine.

Recently, the interest in computer-based systems and sEMG as clinical tools has grown [21, 22]. However, such systems are not yet available everywhere due to the required technical equipment and the rather limited ease of practical implementation and application [11]. Therefore, clinical routine, but also randomized controlled trials have to rely on such subjective assessment tools. In 2017, Scheller et al. stressed this issue as a major problem regarding their randomized multi-center phase III trial on the efficacy of prophylactic nimodipine treatment in vestibular schwannoma (VS) surgery [19]. Potentially, they argued, the large interrater variability of their HB grades may have obfuscated smaller effects of vasoactive treatment on postoperative facial nerve function and hearing.

4.3 sEMG for grading of facial nerve function

In patients with vestibular schwannoma surgery, other studies on the use of sEMG for grading of facial function are not available. The general approach of sEMG-based evaluation of facial function outside the surgical field however has gained interest. For example Ryu et al. [21] investigated sEMG as an assessment tool for facial palsy. They conducted sEMGs on 50 patients with peripheral facial nerve palsy at different points in time (1, 3–4 and 5–6 weeks after onset) and analyzed the correspondence of the individual facial regions with nerve conduction studies (NCS) and several clinical assessment systems including HB. Placement and number of electrodes in their study were similar to ours and the facial movements were almost identical. They report a good correspondence to NCS and clinical assessment. A study by Kim et al. [9] showed similar results when comparing sEMG recordings of 21 patients with peripheral nerve palsy with three common clinical assessment scores (HBGS, Sunnybrook Facial Grading System and Yanagihara grading system).

Our results are in line with these studies and further support sEMG as a useful tool for the assessment of facial nerve function also in a surgical setting. Above and beyond, we developed an alternative grading method, with which an objective classification into different degrees of severity is possible based on the sEMG.

Another advantage of sEMG is the objective comparability of both face halves of individual patients. This might be useful in case of synkinesis. Bernardes et al. [26] showed that there is significant synkinesis even in healthy patients between both halves. This suggests a need for an objective but still comparable parameter which can evaluate the healthy and affected half of the face separately. In comparison, by evaluating the HBGS, the face of the patient is only visible to the examiner as a whole. Subjective visual assessment of the degree of synkinesis and taking this into account for facial function grading may be challenging and may be less reliable. Our current implementation potentially suffers from the effect of synkinesis. However, as separate measurements are available, it is conceivable to extend the procedure to take synchronization between face halves into account and thus limit the impact of synkinesis or separately take this into account for grading.

Regarding the optimal electrode number and placement, Schumann et al. [23] conducted research on facial muscle activation patterns in healthy individuals with sEMG. They showed objective and statistically representative sEMG patterns of symmetric facial muscle function in 30 healthy subjects. Their primary objective was to contribute reference data for neurological examination guidelines with the focus on the optimal number and positioning of the electrodes. Recordings used 48 electrode pairs compared to six

electrodes in our study. An advantage of a high number of electrodes might be the more precise detectability of the distribution of muscle activity. However, as the authors mention themselves, the high number of electrodes complicates evaluation and interpretation in clinical routine. Fewer electrodes with however optimal placement may be sufficient for evaluation due to practical, simple, and thus reproducible positioning. To this end, further studies with the aim of finding optimal electrode placement are required.

4.4 Limitations

A major limitation of our study is the comparably small sample. This likely led to a limited accuracy of performance estimation and facilitated overfitting. To minimize any dependence of the training and test data the leave-one-out approach was chosen. This strategy counteracts as well the possible confounder of the time course as the course of disease, if the data occur from the same patient which cannot appear in the training and test split in one run. The performance of all classifiers would clearly benefit from a much larger training sample. With a larger and more representative dataset, acquired in future studies, an improved analysis could also use the absolute EMG amplitudes instead of or additionally to the lateralization indices. In summary, the reported results probably overestimate the performance in the current data. However, they provide a proof-of-principle to conduct larger studies with potentially overall better performance of the method.

A limitation in the same vein is that not all HB-grades were equally represented. This however reflects the clinical presentation of facial function before and after surgery [27]. Because of this issue we have evaluated the three scenarios and also limited scenario 3 to only HB 1, 2 and 3 to counteract this imbalance. Nevertheless, we believe the reported AUC values demonstrate the viability of our approach but are likely optimistic. Ideally, a larger dataset should be balanced in terms of HB grades for optimal training success.

Although equipment to record sEMG should be available in many neurosurgical centers, the measuring effort itself including placing the electrodes is another disadvantage. In addition, the patient needs to sit in an upright position and must be able to collaborate during the time-consuming measurement. Therefore, the procedure may in fact mostly be viable for research purposes—corresponding to our study aims.

4.5 Conclusions

In summary, this pilot study shows that sEMG can be used in principle to grade facial nerve function and may be a potential alternative or addition to HBGS, primarily for scientific investigations. Our results showed overall good differentiation capacity between clinically relevant HB-ranges. Furthermore, our machine learning based approach is an automated method and yields reproducible results.

Given the need for even more accurate and precise classification systems our findings make an important contribution towards more objectivity. Because of the setting as a pilot study and limitations regarding sample size and capturing all HB ranges further studies are planned. Above all, more data in general as well as the different degrees of severity of facial palsy are to be represented equally, in order to improve and accurately evaluate performance.

Funding Open Access funding enabled and organized by Projekt DEAL. This study was supported by the Deutsche Forschungsgemeinschaft (PR-1275/1-2 and RA 2062/3-1).

Compliance with ethical standards

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethical approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the Martin-Luther University Halle-Wittenberg (Ref. Number 2015-53).

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Cushing H. Tumors of the nervus acusticus and the syndrome of the cerebellopontine angle. Philadelphia and London: W. B. Saunders Co.; 1917.
2. Ho AL, Scott AM, Klassen AF, Cano SJ, Pusic AL, Van Laeken N. Measuring quality of life and patient satisfaction in facial paralysis patients: a systematic review of patient-reported outcome measures. Plast Reconstr Surg. 2012;130(1):91–9.
3. Coulson SE, O'Dwyer NJ, Adams RD, Croxson GR. Expression of emotion and quality of life after facial nerve paralysis. Otol Neurotol. 2004;25(6):1014–9.
4. Bateman N, Nikolopoulos TP, Robinson K, O'Donoghue GM. Impairments, disabilities, and handicaps after acoustic neuroma surgery. Clin Otolaryngol Allied Sci. 2000;25(1):62–5.

5. House JW, Brackmann DE. Facial nerve grading system. Otolaryngol Head Neck Surg. 1985;93(2):146–7.
6. Scheller C, Wienke A, Tatagiba M, Gharabaghi A, Ramina KF, Scheller K, et al. Interobserver variability of the House-Brackmann facial nerve grading system for the analysis of a randomized multi-center phase III trial. Acta Neurochir (Wien). 2017;159(4):733–8.
7. Kang TS, Vrabec JT, Giddings N, Terris DJ. Facial nerve grading systems (1985-2002): beyond the House-Brackmann scale. Otol Neurotol. 2002;23(5):767–71.
8. Yanagihara N. Grading of facial palsy. In: Fisch U Facial nerve surgery. Proceedings of the Third International Symposium on Facial Nerve Surgery; Zurich; 1976. pp. 533–5 Birmingham; Aesculapius Publishing Company 1977.
9. Murty GE, Diver JP, Kelly PJ, O'Donoghue GM, Bradley PJ. The Nottingham system: objective assessment of facial nerve function in the clinic. Otolaryngol Head Neck Surg. 1994;110(2):156–61.
10. Coulson SE, Croxson GR, Adams RD, O'Dwyer NJ. Reliability of the "Sydney," "Sunnybrook," and "House Brackmann" facial grading systems to assess voluntary movement and synkinesis after facial nerve paralysis. Otolaryngol Head Neck Surg. 2005;132(4):543–9.
11. Ross BG, Fradet G, Nedzelski JM. Development of a sensitive clinical facial grading system. Otolaryngol Head Neck Surg. 1996;114(3):380–6.
12. de Ru JA, Braunius WW, van Benthem PP, Busschers WB, Hordijk GJ. Grading facial nerve function: why a new grading system, the MoReSS, should be proposed. Otol Neurotol. 2006;27(7):1030–6.
13. Fattah AY, Gurusinghe AD, Gavilan J, Hadlock TA, Marcus JR, Marres H, et al. Facial nerve grading instruments: systematic review of the literature and suggestion for uniformity. Plast Reconstr Surg. 2015;135(2):569–79.
14. Smith IM, Murray JA, Cull RE, Slattery J. A comparison of facial grading systems. Clin Otolaryngol Allied Sci. 1992;17(4):303–7.
15. Vrabec JT, Backous DD, Djalilian HR, Gidley PW, Leonetti JP, Marzo SJ, et al. Facial nerve grading system 20. Otolaryngol Head Neck Surg. 2009;140(4):445–50.
16. Banks CA, Bhama PK, Park J, Hadlock CR, Hadlock TA. Clinician-graded electronic facial paralysis assessment: the eFACE. Plast Reconstr Surg. 2015;136(2):223e–30e.
17. Banks CA, Jowett N, Hadlock TA. Test–retest reliability and agreement between in-person and video assessment of facial mimetic function using the eFACE facial grading system. JAMA Facial Plast Surg. 2017;19(3):206–11.
18. Prell J, Rachinger J, Scheller C, Alfieri A, Strauss C, Rampp S. A real-time monitoring system for the facial nerve. Neurosurgery. 2010;66(6):1064–73. discussion 73.
19. Scheller C, Wienke A, Tatagiba M, Gharabaghi A, Ramina KF, Ganslandt O, et al. Prophylactic nimodipine treatment for cochlear and facial nerve preservation after vestibular schwannoma surgery: a randomized multicenter Phase III trial. J Neurosurg. 2016;124(3):657–64.
20. Scheller C, Wienke A, Wurm F, Simmermacher S, Rampp S, Prell J, et al. Neuroprotective efficacy of prophylactic enteral and parenteral nimodipine treatment in vestibular schwannoma surgery: a comparative study. J Neurol Surg A Cent Eur Neurosurg. 2014;75(4):251–8.
21. Ryu HM, Lee SJ, Park EJ, Kim SG, Kim KH, Choi YM, et al. Study on the validity of surface electromyography as assessment tools for facial nerve palsy. J Pharmacopuncture. 2018;21(4):258–67.
22. Kim JU, Lee HG, Jung DJ, Choi YM, Song BY, Yook TH, et al. A study on the correlation between surface electromyography and assessment scale for facial palsy. The Acupuncture. 2013;30(5):107–16.
23. Schumann NP, Bongers K, Guntinas-Lichius O, Scholle HC. Facial muscle activation patterns in healthy male humans: a multi-channel surface EMG study. J Neurosci Methods. 2010;187(1):120–8.
24. Choi YM, Kim JU, Kim LH, Yook TH. A study of the electrical properties of the buccal area using facial surface electromyography. The Acupuncture. 2017;34(2):75–82.
25. Koos WT, Day JD, Matula C, Levy DI. Neurotopographic considerations in the microsurgical treatment of small acoustic neurinomas. J Neurosurg. 1998;88(3):506–12.
26. Bernardes DFF, Bento RF, Goffi Gomez MVS. The contribution of surface electromyographic assessment for defining the stage of peripheral facial paralysis: flaccid or sequelae stage. Int Arch Otorhinolaryngol. 2018;22(4):348–57.
27. Falcioni M, Fois P, Taibah A, Sanna M. Facial nerve function after vestibular schwannoma surgery. J Neurosurg. 2011;115(4):820–6.

Journal of Clinical Monitoring and Computing
https://doi.org/10.1007/s10877-022-00928-9

ORIGINAL RESEARCH

Neural networks for estimation of facial palsy after vestibular schwannoma surgery

Stefan Rampp[1,2] · Magdalena Holze[1] · Christian Scheller[1] · Christian Strauss[1] · Julian Prell[1]

Received: 3 May 2022 / Accepted: 4 October 2022

Abstract

Purpose Facial nerve damage in vestibular schwannoma surgery is associated with A-train patterns in free-running EMG, correlating with the degree of postoperative facial palsy. However, anatomy, preoperative functional status, tumor size and occurrence of A-trains clusters, i.e., sudden A-trains in most channels may further contribute. In the presented study, we examine neural networks to estimate postoperative facial function based on such features.

Methods Data from 200 consecutive patients were used to train neural feed-forward networks (NN). Estimated and clinical postoperative House and Brackmann (HB) grades were compared. Different input sets were evaluated.

Results Networks based on traintime, preoperative HB grade and tumor size achieved good estimation of postoperative HB grades ($chi^2 = 54.8$), compared to using tumor size or mean traintime alone ($chi^2 = 30.6$ and 31.9). Separate intermediate nerve or detection of A-train clusters did not improve performance. Removal of A-train cluster traintime improved results ($chi^2 = 54.8$ vs. 51.3) in patients without separate intermediate nerve.

Conclusion NN based on preoperative HB, traintime and tumor size provide good estimations of postoperative HB. The method is amenable to real-time implementation and supports integration of information from different sources. NN could enable multimodal facial nerve monitoring and improve postoperative outcomes.

Keywords Facial nerve · Intraoperative monitoring · Vestibular schwannoma · Machine learning

1 Introduction

Intraoperative monitoring is applied in cerebello-pontine-angle (CPA) surgery to detect and avoid nerval damage. In vestibular schwannoma (VS) surgery, monitoring of free-running EMG, facial motor evoked potentials (MEP) and direct nerve stimulation (DNS) support preservation of facial and vestibulocochlear function and consequently postoperative quality of life [1, 2]. Monitoring of free-running EMG examines continuous EMG activity recorded by needle electrodes in the facial muscles for specific pathological patterns, so-called "A-trains". The overall quantity of A-trains ("traintime") has been shown to correlate with the degree of postoperative facial palsy [3, 4]. The positive predictive value of the method with fixed risk thresholds is ~64%, which is comparable to the values published for MEP and DNS [1, 4].

A limiting factor is the occurrence of false-positive cases with high amounts of A-trains and no severe deterioration of facial function [1, 5, 6]. In a previous study [6], we demonstrated that such patients frequently show a so-called "split" facial nerve [7]. In these cases, the intermediate nerve (NI) takes a course in the CPA separate from the facial nerve, carrying motor fibers targeting the facial muscles [6, 8–10]. Irritation of the NI provokes comparably large amounts of A-trains. Potentially due to the low functional importance of intermedius motor fibers, this is frequently not accompanied by respective deficits [6]. Unfortunately, characteristics of "intermedius" A-trains are not significantly different from "facial" A-trains [11], which prevents differentiation of the two entities. Instead, so-called A-train "clusters", i.e. A-trains occurring in most recording channels within a short time are more frequent in patients with separate NI on a group level [11]. In addition, the observation of a separate

✉ Stefan Rampp
Stefan.rampp@uk-halle.de

[1] Department of Neurosurgery, University Hospital Halle (Saale), Ernst-Grube Str. 40, 06120 Halle, Germany

[2] Department of Neurosurgery, University Hospital Erlangen, Schwabachanlage 6, 91054 Erlangen, Germany

Published online: 04 November 2022

NI increases with larger tumor size, however is rare in cases with very large tumors [11].

These findings suggest complex interactions between tumor size, NI, surgical manipulation, A-train activity and correlation to outcome. It seems therefore unsurprising that fixed traintime thresholds largely independent of tumor size and without consideration of a separate NI suffer from limitations.

In the current study, we employ machine learning and specifically neural networks (NN) to calculate an outcome parameter similar to House-Brackmann (HB) grades [12] based on traintime, tumor size and preoperative functional status. An advantage relevant to our application is the ability to integrate different data types and to capture complex interactions. While understanding the performance of a successful neural network is notoriously difficult, even a pure black-box approach may have clinical merit if it outperforms estimation based on direct interpretation of parameters alone.

The main goal of our study is therefore to provide an improved tool to estimate postoperative facial nerve outcome with the potential for real-time intraoperative application for facial nerve monitoring.

2 Methods

2.1 Patients

Data from 200 consecutive adult patients who had undergone VS surgery between 7/2006 and 8/2016 were selected retrospectively and anonymized. This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of the University Hospital Halle (Saale) (Ref. Number 2018 – 138). All patients of whom data were included in the study had given their written informed consent for scientific usage of their data. Inclusion criteria were first VS surgery, availability of complete continuous intraoperative EMG recordings from clinical routine as well as facial nerve outcome data from follow-up after at least 6 months. Exclusion criteria were previous irradiation and neurofibromatosis.

2.2 Recordings

Continuous EMG was recorded during the complete surgical procedure as described previously [3, 4]. In short, 15 mm long non-insulated needle electrodes were placed parallel in the facial muscles with an interelectrode distance of 5 mm. For each of the 3 main nerve branches 4 electrodes were positioned on the operated side. Referencing neighboring electrodes resulted in 3 bipolar channels per branch. The ground electrode was placed in the contralateral upper arm. Data were recorded with a Grass-Telefactor 15LT biosignal amplifier (West Warwick, RI, USA) with approximately 7 kHz and using a 5 Hz high pass filter.

2.3 EMG processing

Recorded data was evaluated postoperatively by computer-assisted visual inspection using in-house software. Extending automated marking [3], on- and offsets of individual A-trains were marked. In addition, A-train clusters [11] were identified visually. Subsequently, the durations of all A-train events were summed up per channel, yielding a total of 9 traintime values for each patient.

2.4 Clinical data

Clinical data were extracted from clinical documentation: preoperative and immediate postoperative facial nerve function as well as follow-up after 6 months, graded according to House-Brackmann [12]. The HB grading system distinguishes 6 degrees of facial palsy: 1 – normal function, 2 to 5 represent dysfunction from mild to severe and 6 represents total paralysis. Clinically especially relevant is HB ≥ 4 as eye closure on the affected side is no longer possible. HB degrees were checked and corrected, if necessary, by a single experienced evaluator (author JP) to reduce limited interrater reliability [13]. Intraoperative observation of a separate NI was taken from the surgeon's documentation.

2.5 Relationship to postoperative outcome

Relationship of traintime, tumor size and NN estimates of postoperative outcomes (postoperative and follow-up HB grades) with the actual observed outcomes was evaluated using Spearman partial rank correlation as applied previously [4]. A statistically significant partial correlation suggests an association which is not explained by the covariates, e.g. traintime and outcome independent of tumor size [4]. Evaluation of the correlation of only the raw traintime and tumor size with the outcome, i.e. without first passing through the networks was performed to yield a baseline performance to compare network outputs against.

2.6 Neural networks and logistic regression models

Feed-forward networks with different input parameters, a single hidden layer and simultaneous postoperative and follow-up HB grades as outputs were constructed using the *feedforward* function of the Matlab Deep Learning Toolbox (Matlab R2021a, The Mathworks, Natick, MA, USA). Number of hidden layer neurons was chosen equal

to the number of inputs. Continuous network outputs were rounded and interpreted as estimated HB grades. The networks therefore were trained to recognize the association between input parameters and the target "patterns" of HB grade pairs (postoperative and follow-up).

The procedure utilized a Levenberg-Marquardt training function and mean squared error for performance evaluation. Data was randomly separated into 75% (150 datasets) training and 25% (50 datasets) validation splits. Performance was evaluated in only the validation split by calculating chi^2 statistic between estimated HB grades and postoperative and follow-up HB grades. For more intuitive interpretation, chi^2 values were transformed into Cramér's V effect sizes. For 5×5 tables (evaluated HB 1–5), values below 0.05 are considered negligible, 0.05–0.13 small, 0.13–0.22 medium and above 0.22 as large [14].

To illustrate the performance of a more transparent model, multivariable multinomial logistic regression models (LRM) were trained and evaluated with the feature combination showing the best NN performance, applying the same methodology.

2.7 Statistical evaluation of performance

NN training depends on random choice of training and validation splits as well as random initialization of synapse weights between layers. To better estimate overall NN performance, we applied bootstrapping to sample the performance distribution observed with many networks. The approach repeated a single run of calculations 1000 times, yielding 1000 estimates, i.e., chi^2 values of the comparison between network output and outcomes.

The mean and 95% confidence intervals of the resulting distribution was taken as overall performance. For calculation of significance, the distribution was compared to a surrogate distribution using a Komolgorov-Sminorv (KS) test. The surrogate distribution was constructed by shuffling input data of the validation in respect to the outcome values. Chi^2 values were then calculated using surrogate network output. The procedure was also repeated 1000 times yielding the surrogate distribution.

2.8 Comparison of different input sets

Primary endpoint of our study was to evaluate NN with inputs traintime, tumor size and preoperative facial nerve function. Additionally, we evaluated performance, when adding the information that a separate intermedius and/or A-train clusters were observed. Performance differences are discussed based on 95% confidence intervals (CI). Overlapping CI were interpreted as a lack of significant differences, which is considered conservative [15].

2.9 Evaluation of tumor size

Networks trained on traintime, tumor size and preoperative facial nerve function were further analyzed to study the influence of tumor size. The complete dataset was subdivided into groups according to Koos grades. Chi^2 values were then calculated for each group individually. Due to comparable preoperative HB grades in most patients and therefore also within tumor size subgroups, the observed group correlations then necessarily must depend on traintime. Mean correlations and 95%-CI are reported over all 1000 randomizations. For evaluation of differences between tumor size categories, a general linear regression model (GLM) was fitted to the network estimates, taking tumor size and sample size in the groups into account to control for the different patient numbers in tumor size groups, ranging from 18 with Koos 1 to 70 with Koos 3.

2.10 Influence of a separate intermedius nerve

NN performance was investigated regarding the influence of a separate NI. Based on all 200 patients, chi^2 of estimates and clinical HB grades were calculated for patients with and without separate NI in each of the 1000 randomizations and compared with the KS test. We decided not to perform this evaluation in only the validation split unlike the remaining analysis but in the complete sample. Due to the random selection of 50 cases in each randomization, this would have led to varying and frequently unbalanced percentages of cases with a separate NI. Since chi^2 statistics and to some degree Cramér's V are sensitive to the sample size, comparison to performance of other neural networks evaluated in only the smaller validation split is limited.

3 Results

3.1 Patients

Mean age of the 200 included patients was 51 years (21–80 years). 109 patients were women. Tumor size was Koos 1 in 18 patients, Koos 2 in 57, Koos 3 in 70 and Koos 4 in 55 patients [16]. Preoperative facial nerve function was HB 1 on median (range 1–3, 3 patients with HB 3) [12]. A separate NI was observed intraoperatively in 99 patients.

3.2 Conventional analysis

Table 1 provides an overview of results. Traintime was significantly correlated to postoperative and follow-up HB. In patients without separate NI, correlations were higher than in patients with separate NI. A-train clusters were more

Table 1 Correlations with postoperative and follow-up HB of conventional analysis. Significant correlations are printed in bold

Group	Parameters	HB	Spearman correlation	
			rho	**p**
All patients	Mean traintime	Postop.	**0.397**	**<0.0001**
		Follow-up	**0.323**	**<0.0001**
	Mean traintime without clusters	Postop.	**0.442**	**<0.0001**
		Follow-up	**0.350**	**<0.0001**
	Tumor size (Koos)	Postop.	**0.456**	**<0.0001**
		Follow-up	**0.437**	**<0.0001**
	Mean traintime, tumor size controlled (partial correlation)	Postop.	**0.208**	**<0.005**
		Follow-up	0.123	0.084
Without sep. intermedius nerve	Mean traintime	Postop.	**0.564**	**<0.0001**
		Follow-up	**0.511**	**<0.0001**
	Tumor size (Koos)	Postop.	**0.564**	**<0.0001**
		Follow-up	**0.509**	**<0.0001**
With sep. intermedius nerve	Mean traintime	Postop.	**0.198**	**0.0499**
		Follow-up	0.08	>0.1
	Tumor size (Koos)	Postop.	**0.317**	**<0.0001**
		Follow-up	**0.341**	**<0.0001**

frequently observed in patients with a separate intermedius (Fig. 1). However, removal of A-train clusters resulted in only negligible improvement in the complete group.

Tumor sizes also correlated with outcomes and patients with a separate NI had larger tumors than patients without ($p = 0.0012$, $chi^2 = 15.96$, chi-square test). In patients with a separate NI, correlations of tumor size with facial nerve function were lower compared to cases without (Fig. 2).

Controlling for tumor size, partial correlations yielded significant remaining correlations for immediate postoperative traintime and facial nerve function, but were not significant at follow-up.

3.3 Neural networks and LRM

Using traintime, tumor size and preoperative HB grades as input, mean chi^2 comparing NN estimates and outcomes was $chi^2 = 51.3$ ($p < 0.0001$) corresponding to a Cramér's V of 0.36 evaluated only in the validation split. Tables 2 and 3 as well as Fig. 3 show results in detail.

Using the observation of a separate NI or A-train clusters as additional inputs yielded comparable results. Performance using only tumor size or only mean traintime over all channels yielded considerably lower results.

All input combinations were reevaluated using traintime with manually removed A-train clusters ("corrected

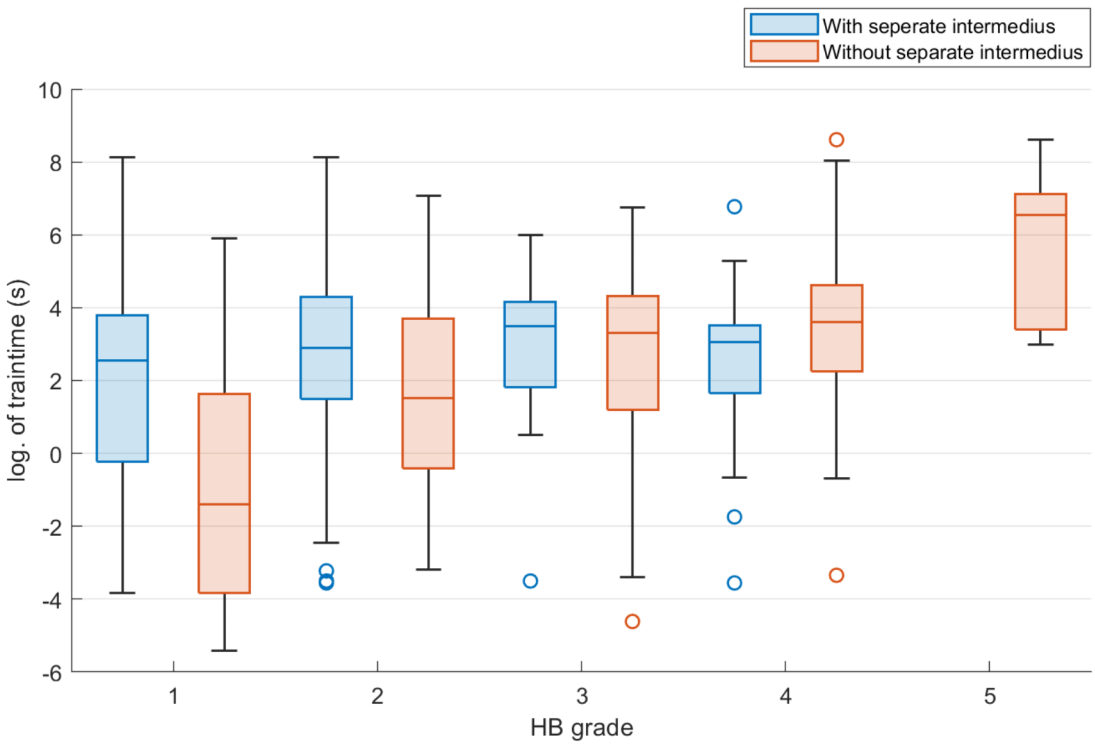

Fig. 1 Mean traintime over all channels in patients with and without a separate intermedius nerve

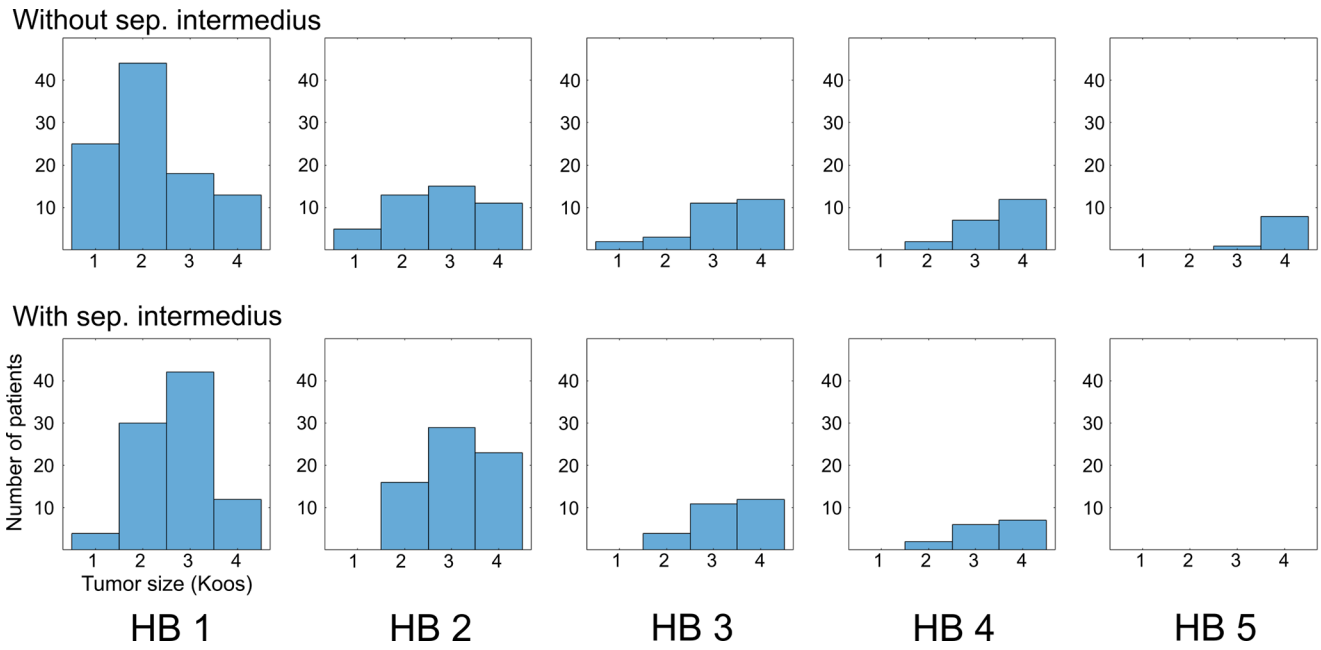

Fig. 2 Koos tumor size in relation to facial nerve outcome. Postoperative and follow-up are pooled

Table 2 Performance of neural network estimates. Results in the validation split (50 patients) are reported

Inputs	Chi^2		Cramér's V
	mean	**CI**	
Koos only	30.6	30.1-31.1	0.28
Mean traintime only	31.9	30.6-33.2	0.28
Mean traintime only (without clusters)	47.7	45.9-49.5	0.35
Traintime, Koos, preOP HB	51.3	49.7-53.0	0.36
+ sep. intermedius	49.1	47.6-50.7	0.35
+ A-train cluster	49.7	48.1-51.3	0.35
+ sep. intermedius and A-train cluster	44.8	43.4-46.2	0.33
Traintime (without clusters), Koos, preOP HB	54.8	53.0-56.7	0.37
+ sep. intermedius	51.6	50.0-53.2	0.36
+ A-train cluster	52.1	50.4-53.8	0.36
+ sep. intermedius and A-train cluster	49.0	47.4-50.7	0.35

Table 3 Performance of neural network estimates in tumor size subgroups. Training was performed with traintime, tumor size and preoperative facial nerve function. Results were then subdivided according to Koos grade for calculation of chi2 and Cramér's V

Tumor size	Chi^2		Cramér's V
	mean	**CI**	
Koos 1	4.8	4.7-5.0	0.08
Koos 2	35.4	34.5-36.0	0.21
Koos 3	118.6	115.2-122.1	0.39
Koos 4	61.3	60.1-62.5	0.28

traintime"). This resulted in a considerable improvement even when only mean traintime over all channels was used as input. Networks with tumor size, preoperative HB and corrected traintime values also resulted in a mean higher chi^2 value.

Using the combination of features with best neural network performance as inputs to LRMs (traintime without clusters, Koos, preoperative HB, $chi^2 = 54.8$, Table 2), yielded a lower mean chi^2 of 41.4 (confidence interval 40.9–41.9), corresponding to a Cramér's V of 0.37.

3.4 Tumor size

Analysis of concordance with postoperative facial nerve function in Koos subgroups are presented in Table 3. Differences of chi^2 values between groups reached statistical significance, also after correcting for the expected tumor and sample size interaction (GLM analysis, $F = 2380$, $p < 0.0001$ for the regression model, $t = -40.8$, $p < 0.0001$ for factor tumor size).

3.5 Influence of a separate intermedius nerve

Comparison of performance in all patients yielded significantly better values in patients without a separate NI using the best set of inputs (preoperative HB, tumor size and corrected traintime): $chi^2 = 164.2$ vs. 65.9 ($p < 0.0001$), corresponding to a Cramér`s V of 0.46 ($n = 99$ patients) and 0.29 ($n = 101$ patients). Networks utilizing corrected traintime showed improved performance only in patients without a separate NI (best chi^2 with A-train clusters: 32.7 vs. 35.6 without and 18.3 vs. 17.0 with a separate NI, Table 4).

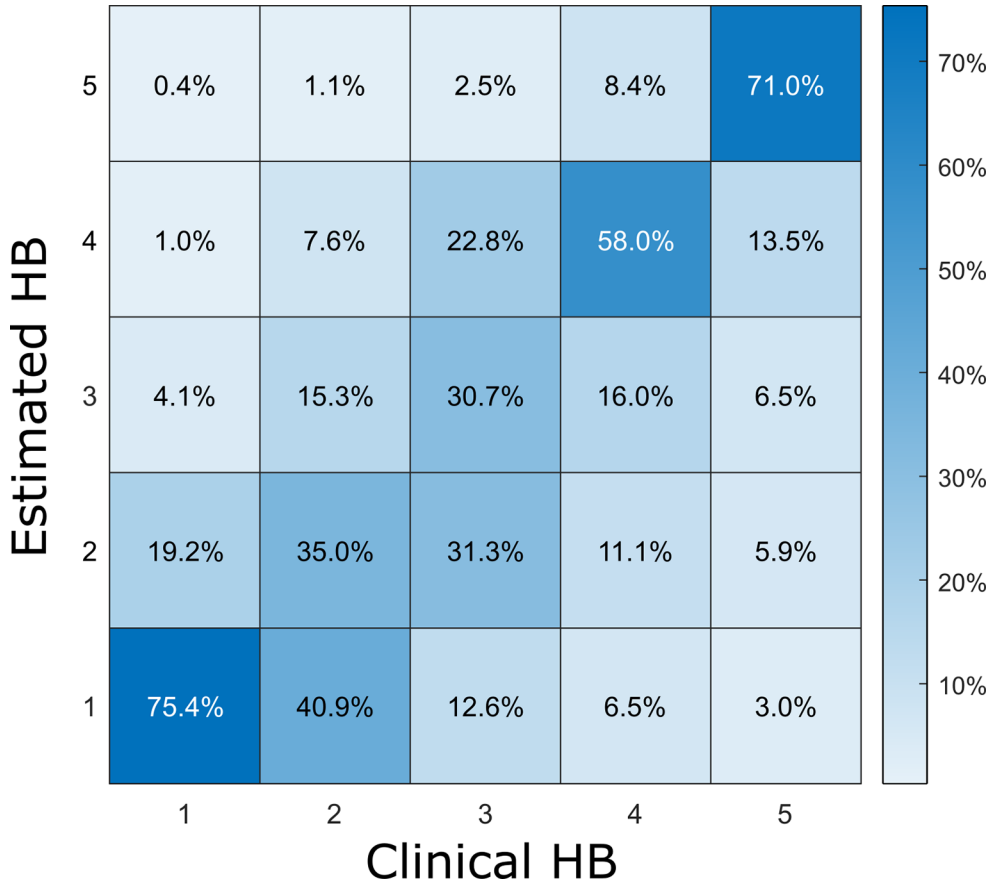

Fig. 3 Concordance of clinical HB grades and neural network estimates of the network with inputs yielding the best results (Traintime without clusters, Koos, preoperative HB-grade). Coloring and percentages in each (independent) column give the portion of all randomized results with a specific HB grade. For example, 75.4% of neural estimates in cases with postoperative or follow-up HB 1 also suggest HB 1, while 19.2% suggest HB 2 and thus overestimate facial nerve palsy.

Table 4 Comparison of neural network performance in patients with and without a separate intermedius nerve. Results in the validation split are reported, grouped according to intraoperative observation of a separate intermedius nerve. Due to the lower sample number in each group (on average approx. 50% due to the portion of patients with separate intermedius nerve), chi^2 and Cramér´s V are generally lower compared to Table 1

Inputs	With sep. intermedius nerve			Without sep. intermedius nerve		
	Chi^2		**Cramér's V**	**Chi^2**		**Cramér's V**
	mean	**CI**		**mean**	**CI**	
Koos only	14.0	13.7-14.4	0.13	21.5	21.0-21.9	0.16
Mean traintime only	10.1	9.7-10.4	0.11	24.3	23.3-25.2	0.17
Mean traintime only (without clusters)	10.3	10.0-10.6	0.11	33.3	32.2-34.5	0.20
Traintime, Koos, preOP HB	18.3	17.8-18.8	0.15	32.7	31.7-33.6	0.20
+ sep. intermedius	17.6	17.1-18.1	0.15	31.9	31.0-32.8	0.20
+ A-train cluster	17.5	17.0-17.9	0.15	33.0	32.0-34.0	0.20
+ sep. intermedius and A-train cluster	15.9	15.4-16.4	0.14	30.6	29.8-31.5	0.19
Traintime (without clusters), Koos, preOP HB	17.0	16.5-17.5	0.15	35.6	34.5-36.7	0.21
+ sep. intermedius	16.7	16.2-17.2	0.15	35.1	34.1-36.0	0.21
+ A-train cluster	16.7	16.2-17.2	0.15	35.2	34.1-36.2	0.21
+ sep. intermedius and A-train cluster	15.5	15.0-16.0	0.14	33.3	32.3-34.2	0.20

4 Discussion

We utilized machine learning approaches in a group of 200 patients undergoing VS surgery. Our results show that these methods can combine preoperative facial nerve function, tumor size and intraoperative traintime to estimate postoperative facial nerve outcomes. Performance exceeds results from evaluation of the features alone and when tumor size is controlled. Performance did not improve when observation of a separate NI and/or detection of A-train clusters were added to the analysis. Prediction improved when A-train-clusters were removed from the detected traintime, mainly due to improvements in patients without a separate NI. Improved prediction may support intraoperative decision

making as well as recognition, which surgical maneuvers carry an increased risk for postoperative palsy.

Our previous studies demonstrated that a separate NI can give rise to an exceeding amount of A-trains not related to postoperative palsy [6, 11], which limits outcome estimation based on free-running EMG alone. Since observation of a separate intermedius is related to tumor size [11], which itself yields predictive information [4, 17, 18], we hypothesized that considering this interaction could improve outcome estimation.

Indeed, integrating preoperative facial nerve function, traintime and tumor size outperformed outcome estimation using only tumor size or traintime. Although performance was generally lower in patients with a separate NI, combined analysis also resulted in improvements in this subgroup.

Preoperative facial nerve function and tumor size have been shown to impact intraoperative monitoring. Facial MEP for example correlate with tumor size already at the start of surgery [19], while traintime interpretation should consider preoperative deficits [3]. Our results show that NN approaches integrate these different modalities, effectively implementing such clinical recommendations in a formalized and objective manner.

Utilizing corrected traintime resulted in a considerable improvement even if only mean traintime was considered. Correction increased chi^2 from 31.9 to 47.7 (Cramér`s V from 0.39 to 0.49). The combination with preoperative HB and tumor size then showed the best of all tested combinations. Correction was based on our previous findings, that patients with separate NI show A-train clusters significantly more often than patients without [11], similar to patients with previous surgery or irradiation [5]. We argued that these clusters are an expression of a hyperexcitable or more vulnerable NI.

The result that removing A-train clusters is beneficial for HB estimation supports the idea that such excessive, clinically not informative traintime may be caused by a separate NI [6, 11]. It is however surprising that considering the observation of a separate intermedius or the presence of clusters to NN was not helpful and even partially decreased performance. Furthermore, the effect was largely present in the subgroup without separate NI, while patients with NI did not benefit (Table 4).

Consequently, the results indicate that A-train clusters generally over-represent actual damage to the facial nerve – not only when a split nerve course is encountered. Cluster traintime should therefore be weighted weaker than traintime from singular A-trains or removed entirely. In the current study, correction however was not sufficient to ameliorate the impact of a separate NI. There are several potential reasons. First, due to practical factors, A-train clusters were identified visually. This strategy may have resulted in marking only the clearest of clusters, while the phenomenon might in fact be subtler and manifest as a "spectrum of over-representation". Furthermore, topography, time and distance between occurrences and relationship to singular A-trains were not evaluated.

Even if such information would not alleviate the intermedius "issue", NN offer further potential improvement. NN allow integration of more information sources, above and beyond the evaluated features. E.g., FMEP [19–21] or direct electrical stimulation [22] could be utilized for a multimodal monitoring approach. In addition, determination of the facial nerve course [23] could add valuable anatomical information.

Overall, estimated HB grades corresponded well to clinical evaluation. In moderate ranges, we observed deviations by one, sometimes two degrees (Fig. 3). Such variability may partially be caused by the subjective nature of HB grading itself, respectively its practical application [24–26]. Scheller et al. [13] investigated the interobserver variability of HB grading as part of a randomized multi-center phase III trial. In this study, too, HB grades varied between observers in an extent comparable to our results. HB grades were also most consistent when facial nerve function was normal or mildly impaired. NN estimates are therefore well within the range of this variability. Further improvement may require the use of a more objective grading system with better inter-rater reliability [26–29].

Finally, a significant disadvantage of neural network is their "black box" nature, i.e., how they achieve their performance is notoriously difficult to interpret. In comparison, LMR are more accessible, as the resulting regression coefficient allow direct interpretation of the relative feature importance. The performance of LMRs in our study was lower than with NN, however still within a clinically useful range. It is conceivable that more training data may result in further improvement. Future studies should therefore conduct more detailed comparisons, including further computational approaches to combine multimodal information.

5 Conclusion

In conclusion, NN using traintime, preoperative facial nerve function and tumor size can estimate postoperative HB grades with good accuracy. However, they do not fully compensate false positive A-train activity associated with a separate NI. Removal of A-train cluster traintime nevertheless seems to be advisable even in cases without a separate course of the intermediate nerve. NN can integrate information from different pre- and intraoperative diagnostic methods and may enable comprehensive multimodal monitoring.

Author contributions All authors contributed to the study conception and design. Data collection was performed by CS, CSt and JP. Analysis was performed by SR and MH. The first draft of the manuscript was written by SR and MH and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the Deutsche Forschungsgemeinschaft (PR-1275/1–2 and RA 2062/3 – 1).
Open Access funding enabled and organized by Projekt DEAL.

Statements and declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

References

1. Prell J, Strauss C, Plontke SK, Rampp S. Intraoperative Funktionsüberwachung des N. facialis: Operationen an Vestibularisschwannomen. HNO. 2017;65:404–12.
2. Stankovic P, Wittlinger J, Georgiew R, Dominas N, Hoch S, Wilhelm T. Continuous intraoperative neuromonitoring (cIONM) in head and neck surgery-a review. HNO. 2020;68:86–92.
3. Prell J, Rachinger J, Scheller C, Alfieri A, Strauss C, Rampp S. A real-time monitoring system for the facial nerve. Neurosurgery. 2010;66:1064–73. discussion 1073.
4. Prell J, Strauss C, Rachinger J, Alfieri A, Scheller C, Herfurth K, et al. Facial nerve palsy after vestibular schwannoma surgery: Dynamic risk-stratification based on continuous EMG-monitoring. Clin Neurophysiol Int Federation Clin Neurophysiol. 2014;125:415–21.
5. Rampp S, Strauss C, Scheller C, Rachinger J, Prell J. A-trains for intraoperative monitoring in patients with recurrent vestibular schwannoma. Acta Neurochir. 2013;155:2273–9.
6. Prell J, Strauss C, Rachinger J, Scheller C, Alfieri A, Herfurth K, et al. The intermedius nerve as a confounding variable for monitoring of the free-running electromyogram. Clin Neurophysiol. 2015;126:1833–9.
7. Strauss C, Prell J, Rampp S, Romstöck J. Split facial nerve course in vestibular schwannomas. J Neurosurg. 2006;105:698–705.
8. Ashram YA, Jackler RK, Pitts LH, Yingling CD. Intraoperative electrophysiologic identification of the nervus intermedius. Otol Neurotol. 2005;26:274–9.
9. Alfieri A, Fleischhammer J, Peschke E, Strauss C. The nervus intermedius as a variable landmark and critical structure in cerebellopontine angle surgery: an anatomical study and classification. Acta Neurochir. 2012;154:1263–8.
10. Alfieri A, Rampp S, Strauss C, Fleischhammer J, Rachinger J, Scheller C, et al. The relationship between nervus intermedius anatomy, ultrastructure, electrophysiology, and clinical function. Usefulness in cerebellopontine microsurgery. Acta Neurochir (Wien). 2014;156:403–8.
11. Rampp S, Illert J, Krempler K, Strauss C, Prell J. A-train clusters and the intermedius nerve in vestibular schwannoma patients. Clin Neurophysiol. 2019;130:722–6.
12. House JW, Brackmann DE. Facial Nerve Grading System. Otolaryngology-Head and Neck Surgery. 1985;93:146–7.
13. Scheller C, Wienke A, Tatagiba M, Gharabaghi A, Ramina KF, Scheller K, et al. Interobserver variability of the House-Brackmann facial nerve grading system for the analysis of a randomized multi-center phase III trial. Acta Neurochir Springer-Verlag Wien. 2017;159:733–8.
14. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale: L. Erlbaum Associates; 1988.
15. Cumming G, Finch S. Inference by Eye: Confidence Intervals and How to Read Pictures of Data. Am Psychol. 2005;60:170–80.
16. Koos WT, Day JD, Matula C, Levy DI. Neurotopographic considerations in the microsurgical treatment of small acoustic neurinomas. J Neurosurg. 1998;88:506–12.
17. Falcioni M, Fois P, Taibah A, Sanna M. Facial nerve function after vestibular schwannoma surgery. J Neurosurg. 2011;115:820–6.
18. Samii M, Matthies C. Management of 1000 vestibular schwannomas (acoustic neuromas): the facial nerve–preservation and restitution of function. Neurosurgery. 1997;40:684–5.
19. Matthies C, Raslan F, Schweitzer T, Hagen R, Roosen K, Reiners K. Facial motor evoked potentials in cerebellopontine angle surgery: Technique, pitfalls and predictive value. Clin Neurol Neurosurg. 2011;113:872–9.
20. Dong CC, Macdonald DB, Akagami R, Westerberg B, Alkhani A, Kanaan I, et al. Intraoperative facial motor evoked potential monitoring with transcranial electrical stimulation during skull base surgery. Clin Neurophysiol. 2005;116:588–96.
21. Greve T, Wang L, Thon N, Schichor C, Tonn JC, Szelényi A. Prognostic value of a bilateral motor threshold criterion for facial corticobulbar MEP monitoring during cerebellopontine angle tumor resection. J Clin Monit Comput Springer Sci Bus Media B V. 2020;34:1331–41.
22. Quimby AE, Lui J, Chen J. Predictive Ability of Direct Electrical Stimulation on Facial Nerve Function Following Vestibular Schwannoma Surgery: A Systematic Review and Meta-analysis. Otol Neurotol NLM (Medline). 2021;42:493–504.
23. Savardekar AR, Patra DP, Thakur JD, Narayan V, Mohammed N, Bollam P, et al. Preoperative diffusion tensor imaging-fiber tracking for facial nerve identification in vestibular schwannoma: A systematic review on its evolution and current status with a pooled data analysis of surgical concordance rates. Neurosurgical Focus. American Association of Neurological Surgeons; 2018. p. 44.
24. Ahrens A, Skarada D, Wallace M, Cheung JY, Neely JG. Rapid simultaneous comparison system for subjective grading scales grading scales for facial paralysis. Am J Otol [Internet]. 1999;20:667–71. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10503592.
25. Alicandri-Ciufelli M, Piccinini A, Grammatica A, Salafia F, Ciancimino C, Cunsolo E, et al. A step backward: The "Rough" facial nerve grading system. Journal of Cranio-Maxillofacial Surgery. J Craniomaxillofac Surg; 2013;41.
26. De Ru JA, Braunius WW, Van Benthem PPG, Busschers WB, Hordijk GJ. Grading facial nerve function: Why a new grading system, the MoReSS, should be proposed. Otology and Neurotology. 2006;27:1030–6.
27. Coulson SE, Croxson GR, Adams RD, O'Dwyer NJ. Reliability of the "Sydney," "Sunnybrook," and "House Brackmann" facial

grading systems to assess voluntary movement and synkinesis after facial nerve paralysis. Otolaryngology - Head and Neck Surgery. Mosby Inc.; 2005;132:pp. 543–9.
28. Murty GE, O'donoghue GM, Bradley PJ, Diver JP, Kelly PJ. The Nottingham System: Objective assessment of facial nerve function in the clinic. Otolaryngology–Head and Neck Surgery. Otolaryngol Head Neck Surg. 1994;110:156–61.
29. Fattah AY, Gurusinghe ADR, Gavilan J, Hadlock TA, Marcus JR, Marres H, et al. Facial nerve grading instruments: Systematic review of the literature and suggestion for uniformity. Plastic and Reconstructive Surgery. Lippincott Williams and Wilkins; 2015. pp. 569–79.

6 Selbstständigkeitserklärung

I hereby declare that I have written this thesis independently and have not used any sources other than those indicated.

Halle (Saale), March 03, 2023 Magdalena Holze

7 Declaration of previous doctoral thesis attempts

I wrote this thesis as part of my first attempt of a doctoral thesis. I submitted this work exclusively to the Medical Faculty of the Martin Luther University Halle- Wittenberg as a doctoral thesis.

Halle (Saale), March 03, 2023 Magdalena Holze

8 Acknowledgement

A big thank you goes to my supervisor PD Dr. Stefan Rampp, who supervised me excellently during the last six years and supported me in all steps of my thesis. If I had any questions or problems, he always had an open ear and enabled me to work independently and to contribute my own ideas to this work.

I also want to thank my supervisor, Prof. Dr. Julian Prell, who always supported and advised me during the study and the preparation and writing of the publications.

From this thesis and the great supervision. I discovered the fun and enthusiasm of scientific working, which I would like to continue in future projects.

At this point I would also like to thank the entire team of the Neurosurgery Department of the University Hospital Halle (Saale) and Prof. Dr. Christian Strauss, who have consistently supported me during the whole project. In particular, I want to thank Christin Zöller and Anke Dietz, who always gladly helped me with the recruitment of study patients.

It is especially close to my heart to thank my family and my friends, who have always supported me along the way and have been there to help and advise me at any time.