Received: 15 July 2022 | Accepted: 16 December 2022
DOI: 10.1002/wjo2.94

RESEARCH PAPER

WJO HNS World Journal of Otorhinolaryngology- Head and Neck Surgery

Facial nerve palsy following parotid gland surgery: A machine learning prediction outcome approach

Carlos M. Chiesa-Estomba[1,2,3] | **Jose A. González-García**[1] | **Ekhiñe Larruscain**[1] | **Jon A. Sistiaga Suarez**[1] | **Miquel Quer**[4] | **Xavier León**[4] | **Paula Martínez-Ruiz de Apodaca**[2,5] | **Celia López-Mollá**[5] | **Miguel Mayo-Yanez**[2,6,7] | **Alfonso Medela**[8]

[1]Department of Otorhinolaryngology—Head and Neck Surgery, Donostia University Hospital, Donosti-San Sebastián, Spain

[2]Head & Neck Study Group of Young-Otolaryngologists of the International Federations of Oto-rhino-laryngological Societies (YO-IFOS), Paris, France

[3]Biodonostia Health Research Institute, San Sebastián, Spain

[4]Department of Otorhinolaryngology, Hospital Santa Creu I Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain

[5]Department of Otorhinolaryngology, Doctor Peset University Hospital, Valencia, Spain

[6]Otorhinolaryngology—Head and Neck Surgery Department, Complexo Hospitalario Universitario A Coruña (CHUAC), A Coruña, Galicia, Spain

[7]Clinical Research in Medicine, International Center for Doctorate and Advanced Studies (CIEDUS), Universidade de Santiago de, Compostela (USC), Santiago de Compostela, Galicia, Spain

[8]LEGIT Health, Bilbao, Spain

Correspondence
Carlos M. Chiesa-Estomba, Department of Otorhinolaryngology—Head and Neck Surgery, Hospital Universitario Donostia, Doctor Begiristain Pasealekua, #1, PC. 20014, Donosti-San Sebastián, Spain.
Email: chiesaestomba86@gmail.com

Funding information
None

Abstract

Introduction: Machine learning (ML)-based facial nerve injury (FNI) forecasting grounded on multicentric data has not been released up to now. Three distinct ML models, random forest (RF), *K*-nearest neighbor, and artificial neural network (ANN), for the prediction of FNI were evaluated in this mode.

Methods: A retrospective, longitudinal, multicentric study was performed, including patients who went through parotid gland surgery for benign tumors at three different university hospitals.

Results: Seven hundred and thirty-six patients were included. The most compelling aspects related to risk escalation of FNI were as follows: (1) location, in the mid-portion of the gland, near to or above the main trunk of the facial nerve and at the top part, over the frontal or the orbital branch of the facial nerve; (2) tumor volume in the anteroposterior axis; (3) the necessity to simultaneously dissect more than one level; and (4) the requirement of an extended resection compared to a lesser extended resection. By contrast, in accordance with the ML analysis, the size of the tumor (>3 cm), as well as gender and age did not result in a determining favor in relation to the risk of FNI.

Discussion: The findings of this research conclude that ML models such as RF and ANN may serve evidence-based predictions from multicentric data regarding the risk of FNI.

Conclusion: Along with the advent of ML technology, an improvement of the information regarding the potential risks of FNI associated with patients before each procedure may be achieved with the implementation of clinical, radiological, histological, and/or cytological data.

KEYWORDS
gland, machine learning, parotid, personalized medicine, surgery

INTRODUCTION

Facial nerve injury (FNI) continues to be considered the most severe complication subsequent to parotid gland surgery (PGS). The estimated transient facial nerve dysfunction is within 20%–65% in patients where a parotidectomy was performed, with a temporary or absolute facial nerve palsy rate of 0%–7% in those patients, which significantly affects their quality of life.[1,2]

Machine learning (ML) is a subset of artificial intelligence (AI) that enables computers to learn from historical data, gather insights, and make predictions about new data using the model learned. In recent years, ML has achieved an increasing interest in different medical subspecialties[3] because it is considered a useful choice that can be used to provide support for clinical decision-making, classification purposes, or to establish prognosis.[4–7]

In recent years, various studies have been divulged regarding the use of ML models in both the medical and otolaryngological fields.[4–10] Following this trend, our group has validated the practicability of the implementation of numerous ML techniques, including the use of an artificial neural network (ANN) for FNI forecast in two proof-of-concept studies.[11,12]

Nevertheless, as we remarked earlier, the necessity of running our ML models using a higher input of anonymized multicentric data was due to the main limitations in using unicentric available data. We designed a multicentric study to measure the efficiency of the implementation of three different ML models in the forecasting of facial palsy in a group of patients. Developing a web-based application for prognostic estimation, which was able to indicate an accurate prediction for each case, on a daily basis was our main goal.

MATERIALS AND METHODS

A retrospective, longitudinal, multicentric study was performed on a group of patients who underwent PGS for benign tumors, from June 2010 to June 2019, at three different university hospitals (Donostia University Hospital, Sant Creu y Sant Pau University Hospital, and Doctor Peset University Hospital). The main goal of this research was to validate the reliability of three different ML models (random forest [RF], *K*-nearest neighbor [KNN], and ANN) in relation to the prognosis of FNI as described in two previously published unicentric studies by our group.[11,12] Case identification was made by a complete review of each department's databases using the International Classification of Diseases (ICD-9-10).

The inclusion criteria included patients older than 18 years with a clinically and radiologically evident benign tumor in the parotid gland. All nonsurgically treated patients were excluded in the case of revision surgery, or if the final histology showed malignancy. Approval from Ethics Committee was obtained (CCH—110419).

Medical histories from patients were outlined to collect information regarding demographic data, clinical presentation, preoperative assessment, radiological test (computerized tomography [CT] or magnetic resonance imaging [MRI]) diagnosis, and surgical

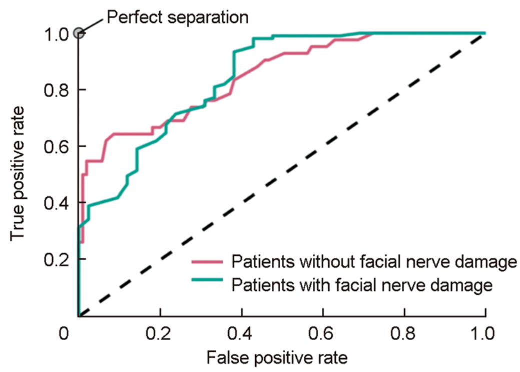

FIGURE 1 Facial nerve injury (FNI) prediction according to the receiver-operating characteristic curve (ROC curve) of the random forest (RF) model.

management. Different results from previously released studies were used.[11,12] ML model analyses were run to determine any type of correlation between the clinical variable and FNI results (Figure 1).

Along with a CT or MRI, ultrasound-guided fine-needle aspiration was performed on every patient ahead of surgery to obtain the anatomic details. Parotidectomy was mainly performed by using the modified Blair or facelift incision reliant on patient preference or the surgeon's experience. Tumor resection was performed following the standard technique. The classification and extension of the parotid gland were established according to the Parotidectomy Classification of the European Salivary Gland Society (ESGS).[13,14] The facial nerve was supervised and stimulated in all cases previously as well as later on. A vacuum-suction drain (Jost–Redon) was normally applied at the end of the procedure.

Following the procedure, the facial nerve function was assessed instantly through brow furrowing, closing eyes, lip whistling shape, and teeth showing. The House–Brackmann scale was applied to measure facial function during the follow-up.

Quantitative variables are detailed by mean ± standard deviation. Outcomes are shown in both total and percentage. Two-samples *t*-tests were applied to target differences between patients with FNI in the case of persistent variables, and χ^2 or Fisher's exact test for unconditional random variables. Differences were adjudicated as statistically significant when two-tailed p value was less than 0.05. Results among the ML models were compared with the implementation of a Logistic Regression Model (LRM) previously built and programmed. Statistical analyses were achieved using JASP (Version 0.11.1, University of Amsterdam, Netherlands; https://jasp-stats.org/).

Respecting model training, a predictive archetype rooted on parameters collected in the clinical practice was designed over the training data. Those strictures included age, gender, medical background, histology, size (>3 cm), anterior-to-posterior, mid-to-lateral, cephalad-to-caudal maximum diameter (obtained from radiological imaging and confirmed with the final histological report), anatomical situation, and levels involved according to the ESGS classification.[13,14] Data pretreatment was processed to identify any latent-imbalanced distribution of different classes and/or to avoid the negative impact of classification issues on model fitting or to minimize disparity in the frequencies of the observed classes.

Our data set was separated by applying 80:20 stratified sampling in accordance with the FNI result; therefore, the ML algorithms were prepared by taking 80% of the applicable cases and evaluating using the remaining 20%. Following this, it was run a recursive elimination feature with the purpose of targeting the most severe variables with the intention of enhancing the ML classification performance. To safeguard model stability and reduce bias, 5-fold cross-validation was performed by RF and KNN algorithms. The average results on the test fold are presented.

An ANN model embedded with a precise architecture (two hidden layers of 200 and 100 neurons, respectively) was selected for tabular data purposes.

The ANN, encloses the absolute variables, performs a dropout before feeding input data into the linear layers, and applies batch normalization to continuous variables. The output layer consists of two-neuron layers where each one presents one of the target categories. The model was trained for 20 epochs with a learning rate of 0.01 and another 4 epochs with a learning curve 10 times smaller. Hyperparameters were tuned to maximize the area under the ROC (receiver-operating characteristic) curve (AUC).

The classification according to the performance of the 3 ML algorithms was then tested by comparing the AUC (internal validation). All predictive models were later externally validated, reporting the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), accuracy, recall, and *F*-score. RF and KNN data analyses were implemented with multiple linear regression, RF, caret, and e1071 libraries from R package version 3.5.0 (R Foundation for Statistical Computing; https://www.r-project.org/) and JASP (Version 0.11.1; University of Amsterdam; https://jasp-stats.org/). ANN algorithms were developed and performed with PyTorch version 1.6 (https://pytorch.org/), and Sciket-Learn version 0.17.1 (http://scikit-image.org). Data preprocessing and analysis were achieved with Pandas version 1.1.0 (https://pandas.pydata.org). The primary outcome resulted in the presence of transient or permanent FNI in at least one branch of the facial nerve after the procedure.

RESULTS

During the research stage, 736 patients were intervened due to the existence of benign tumors in the parotid gland. Three hundred and ninety-seven (53.9%) patients were male and 339 (46.1%) were female. The average age was 55 years (SD: 8/minimum 18/maximum 87). The mean follow-up was 12 months (minimum 6/maximum 24) (Table 1). There were no differences found between groups ($p = 0.947$).

The most prevalent final histological diagnosis was pleomorphic adenoma, which was found in 368 (50.0%) of the cases. The most regular type of resection was type I (333; 45.2%). The parotid tail resulted as the most usual sublocation involved (322; 43.8%) in relation to the levels resected; the most frequent was the level II (284; 38.6%) followed by the combination of level I and II (188; 25.5%) Transient facial nerve paresis resulted in 213 (28.9%) patients and 34 (4.6%) experienced a conclusive facial nerve or facial nerve branch paralysis. (Table 1). In the LRM, the situation and levels included emerged as the most weighted and significant variables (Table 2).

Respecting the prediction exactness and performance of each ML algorithm, the RF and the ANN accomplished the highest sensitivity, specificity, PPV, NPV *F*-score, ROC-AUC, and accuracy in general (Tables 3 and 4). The RF algorithm reached performance values above 0.9 for NPV. RF and ANN achieved values above 0.8 for specificity, accuracy, and ROC-AUC. RF obtained results above 0.8 for sensitivity, *F*-score, precision, and recall; and ANN showed the highest PPV. Overall, in accordance with our previous studies, the algorithms were biased toward the majority class of non-FNI patients, showing very low sensitivity and PPV which is related to the high number of false negatives in the case of the RF and KNN algorithms. As we established formerly, the almost entire overlap between FNI and non-FNI patient classes results in an added difficulty to determine and discriminate between these classes when using our data set variables. This constraint highlights the importance of testing each algorithm separately. Regarding the most influential predictors in the performance of the RF models, the location of the tumor on the gland (mid, superior, and deep lobe), levels dissected, the anterior to posterior size, and the type of resection were the most significant factors associated to the risk of FNI (Table 4 and Figures 1, 2, and 3). The anterior to posterior size, situation, levels, and type of resection emerged as the most angled predictors for our ANN algorithm.

DISCUSSION

Following the former publication of two proof-of-concept studies,[11,12] we aimed to elucidate the performance of ML algorithms over multicentric data. We included the same variables obtained from the clinical evaluation and data linked to surgery that might be contingent in clinical scenarios, to foresee FNI after the procedure for benign tumors of the parotid gland. According to our results, the RF and ANN turned out to be the most precise models.

Old age, malignancy, tumor size (>70 cm^3), operative time, the need for revision surgery due to recurrence, tumor subsite location (superficial vs. deep), and extent of surgery[15–26] are well-known factors related to an increased risk of temporary or permanent FNI after PGS for benign tumors. Nevertheless, the data reported here are heterogeneous, due to the different types of surgical techniques (i.e., extracapsular dissection, partial parotid gland resection, and superficial or total parotidectomy). For this reason, in this study, the authors decided to present results in compliance with the ESGS classification system.[13,14]

According to our results using the RF and the ANN algorithms, our results remain to be instinctive. The most substantial factors associated with a risk escalation of FNI were as follows: (1) location,

TABLE 1 Demographic and clinical data.

Variable	*n*	%	*p* Value
Sex			
Male	397	53.9	0.262
Female	339	46.1	
Mean age	55 ± 8 years (min: 18/max: 87)		0.09
Size			
<3 cm	456	62	0.19
>3 cm	280	38	
Maximum length per plane (cm^3)			
Anterior to posterior	2.58 ± 1.20 (min: 0.90/max: 9.0)		0.003
Medial to lateral	2.12 ± 1.09 (min: 0.70/max: 6.5)		0.001
Cephalic to caudal	2.22 ± 0.97 (min: 0.76/max: 8.7)		0.132
Type of resection			
I	333	45.2	0.452
II	161	21.9	0.186
III	179	24.3	0.021
IV	63	8.6	0.001
Anatomical situation in the gland			
Parotid tail	322	43.8	0.432
Mid-lobe	89	12.1	0.022
Superior lobe	48	6.5	0.234
Deep lobe	74	10.1	0.001
Superior and middle lobes	33	4.5	0.001
Inferior and middle lobes	121	16.4	0.338
All superficial	26	3.5	0.111
Accessory	23	3.1	0.931
Levels			
I	129	17.5	0.089
II	284	38.6	0.239
I + II	188	25.5	0.001
II–III	68	9.2	0.056
I–IV	16	2.2	0.001
III + IV	24	3.3	0.174
I–III	26	3.5	0.134
V	23	3.1	0.267
Histology			
Pleomorphic adenoma	368	50	0.098
Whartin tumor	260	35.3	0.416
First branch branchial cyst	18	2.4	0.223
Oncocytoma	16	2.2	0.435

TABLE 1 (Continued)

Variable	*n*	%	*p* Value
Basal cell adenoma	15	2.0	0.347
Oncocytic papillary cystoadenoma	12	1.6	0.611
Reactive lymphadenitis	12	1.6	0.113
Microcystoadenoma	9	1,2	0.178
Acinar cell tumor	7	0.9	0.099
Lipoma	6	0.8	1
Kuttner tumor	5	0.7	0.987
Myoepithelioma	5	0.7	0.789
Chondroma	3	0.4	1
Transient facial palsy			
Yes	213	28.9	
No	523	71,1	
Definitive facial palsy (branch)			
Yes	34	4.6	
No	702	95.4	

Note: Type of resection according to the ESGS: I, parotidectomy one level or extracapsular dissection; II, parotidectomy (one or two levels, more often partial superficial); III, parotidectomy (two levels, more often superficial); IV, parotidectomy (three or four levels removed, more often total). The anatomical situation in the gland corresponds to the clinical description by the surgeon. Levels according to ESGS classification: cranial superficial = I; caudal superficial = II; deep caudal = III; deep cranial = IV; accessory = V.

Abbreviations: ESGS, European Salivary Gland Society; max, maximum; min, minimum.

TABLE 2 Variable weight according to the LMR and variable importance according to the ANN.

LMR				ANN	
Variable	*P* Value	OR	95% CI	Variable	Importance
Sex	0.081	0.735	0.520–1.039	Volume AP	5.400331
Type of resection	0.458	0.686	0.558–0.842	Situation	4.541683
Volume >3 cm	0.600	0.862	0.494–1.503	Levels	4.247535
Situation	0.007	0.890	0.817–0.969	Type of resection	4.050514
Levels	0.009	0.850	0.752–0.959	Volume CC	3.654956
Histology	1	1.002	0.940–1.064	Histology	3.202214
Age	0.113	1.009	0.998–1.020	Volume ML	3.180621
Volume ML	0.971	0.994	0.712–1.386	Age	2.076110
Volume AP	0.756	0.956	0.718–1.273	Sex	0.826466
Volume CC	0.661	1.055	0.829–1.343	Volume >3 cm	0.292022

Abbreviations: ANN, artificial neural network; AP, anteroposterior; CC, cephalocaudal; CI, confidence interval; LMR, logistic multivariant regression; ML, medial to lateral; OR, odds ratio.

in cases of patients whose tumor was situated in the mid-portion of the gland near, over the main trunk of the facial nerve, or placed in the highest part over the frontal, or the orbital branch of the facial nerve; (2) tumor volume in the anteroposterior axis due to the necessity for a higher facial branch dissection; (3) the demand of dissecting more than one level simultaneously; and (4) the necessity for an extended resection compared with a lesser extended resection. Contrastingly, the size of the tumor (>3 cm) was not associated with an increased risk of FNI according to the ML analysis. Also, sex and age were not connected with risk escalation of FNI.

Several articles regarding ML implementations on head and neck cancer treatments have been released, which focused primarily on

TABLE 3 Classification accuracy of each machine learning algorithm on the testing set.

Model	F-score	ROC-AUC	95% CI—for the ROC-AUC	Precision (%)	Recall (%)
Random forest	0.864	0.821	0.794–0.848	86.7	86.4
K-nearest neighbor	0.699	0.643	0.663–0.735	69.6	70.7
Artificial neural network	0.70	0.860	0.827–0.893	74.1	67.3
Logistic regression	0.773	0.693	0.742–0.804	75.3	70.9

Abbreviations: CI, confidence interval; ROC-AUC, receiver-operating characteristic-area under the curve.

TABLE 4 Classification of performance of each machine learning algorithm on the testing set.

Model	Sensitivity (%) (95% CI)	Specificity (%) (95% CI)	Positive predictive value (%) (95% CI)	Negative predictive value (%) (95% CI)	Accuracy (%) (95% CI)
Random forest	89.29 (71.77–97.73)	85.71 (78.12–91.45)	28.52 (20.14–38.69)	99.21 (97.72–99.73)	85.93 (79.24–91.11)
K-nearest neighbor	56.41 (39.62– 72.19)	75.93 (66.75–83.63)	13.01 (8.83–18.75)	96.47 (94.95–97.54)	74.75 (66.93–81.55)
Artificial neural network	67.45 (63.60–70.41)	90.91 (87.81–92.23)	74.60 (70.82–77.24)	87.66 (84.65–89.40)	84.26 (81.34–86.72)
Logistic regression	64.10 (52.44–74.66)	75.34 (71.86–78.59)	14.23 (11.82–17.04)	97.05 (96.06–97.80)	74.67 (71.36–77.78)

Abbreviations: CI, confidence interval; ROC-AUC, area under the receiver-operating characteristic curve.

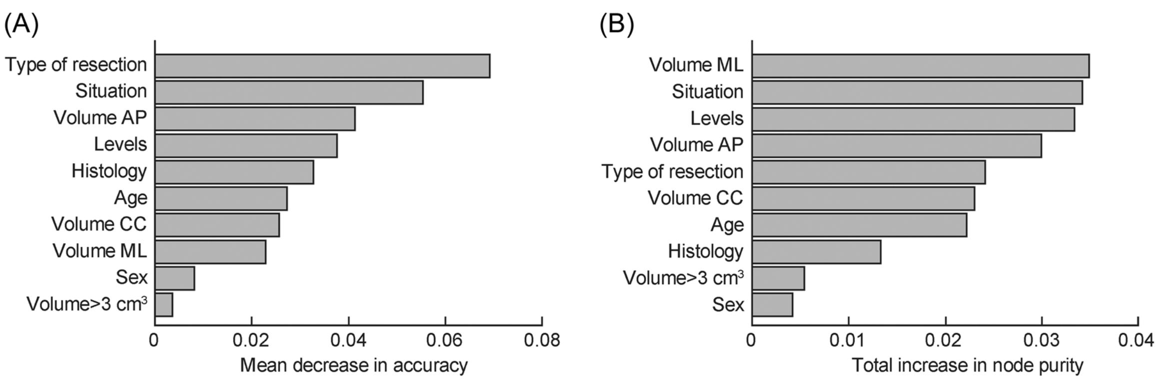

FIGURE 2 Relative rankings of importance of each variable from the random forest algorithm: (A) mean decrease in accuracy and (B) total increase in node purity. AP, anteroposterior; CC, cephalocaudal; ML, medial to lateral.

the following: forecasting survival,[27] improving prognostic predictions,[28] predicting delays in adjuvant treatment for head and neck cancer,[29] analyzing imaging data to create models that predict foretell outcomes,[30] or looking for better models to predict occult nodal metastasis.[31] Regarding FNI prediction in PGS for benign tumors, our group confirmed the reliability of the KNN, RF, and ANN algorithms over unicentric data. However, the urgency to clarify the capabilities of our algorithms concerning discrimination and determination, makes it essential to the process of measuring the performance over multicentric data.

Conceptually, RF is designed to combine multiple decision trees (DTs) with the purpose of enhancing the accuracy and robustness of predictions in contrast to those achieved by a single DT. To perform this task, multiple DTs recursively divide the feature space in a binary fashion. Subsequently, a feature that leads to the largest reduction of the residual sum of squares is selected by each split. Therefore, two distinct regions are obtained at every step of the tree-building process. In every step, the splitting procedure is repeated based on other features and multiple regions in the multivariate space of the observed data are gathered. In this vein, the prediction will vary depending on each region and will match the mean of the observed response variable in the respective region. For RFs, all those multiple trees are built. However, to dodge from building a similar DT, there are just an exact number of features considered in this subdivision. This implements different structures of the accomplished DTs and produces a decorrelation of the trees before being averaged for the final calculation, with the impact of minimizing prediction variance in forthcoming cases.[32]

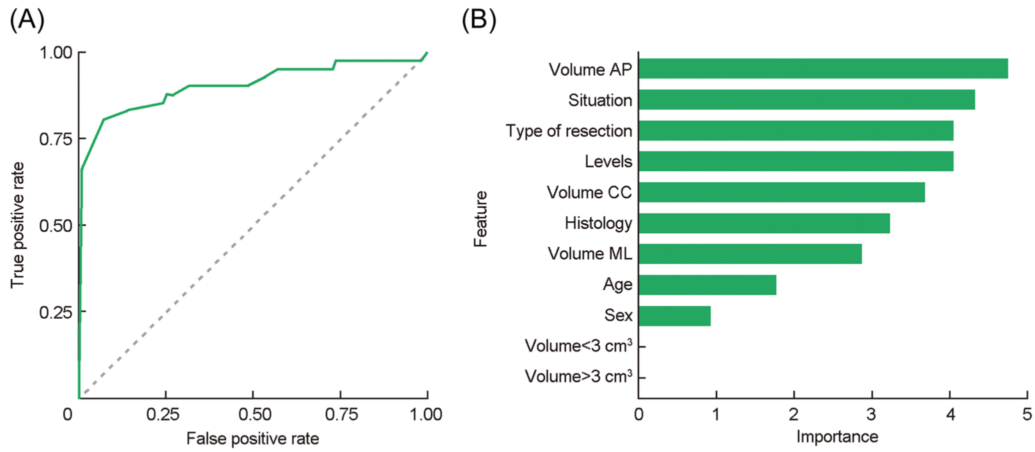

FIGURE 3 Facial nerve injury prediction according to (A) area under the receiver-operating characteristic curve (ROC-AUC) of the artificial neural network and (B) feature of importance graph. AP, anteroposterior; CC, cephalocaudal; ML, medial to lateral.

Meanwhile, ANNs are modeled computer programs designed to simulate the way in which the human brain processes and decodes data. The algorithm gathers information to further apply its knowledge by detecting patterns or relationships within these data and creating the possibility to learn through its own experience and not from pre-existing programming.[4] Structurally, an ANN is formed by hundreds of single units named artificial neurons or perceptrons, which are connected with coefficients, constituting the neural structure presented as a layered architecture. One of the main advantages of the ANN relies on the activation function that can confer nonlinearity to the architecture, improving its capacity to learn any complex relationship between input and output data. In this way, during the training process, the interunit connections may be optimized until the error in predictions is minimized while the network reaches the specified level of accuracy. At the end of the process, once the network is trained and tested, it might receive new input information to predict a precise output.

However, when interpreting those conclusions in the clinical environment, it is important to establish and recognize the differences existing among the architecture of ML algorithms and the classical statistics models. ML algorithms focus exclusively on how different variables interact, constructing predictions regarding an unknown further variable.[31] Contrastingly, statistics primarily make inferences addressing the way components are related to each other through the development of the statistical model.[33,34] Thus, both fields, correctly overlapped, success in providing complementary results.

Regression models have been widely applied in medical research over time. Despite the reliability of the aforementioned algorithms to boost both diagnostic and management precision, it is adequately critical to emphasize the two main inadequacies: the assumption of normality concerning residuals, and their inability to identify and target nonlinear relationships.[35] ML methods are evolving to surpass the limitations of traditional outcome-prediction methods, gaining enhanced implementations in the field of Otolaryngology.[31] Nevertheless, the utilization of ML methods is limited in the actual clinical scenario, due to a lack of understanding and nonexisting easier implementable methods and applications.

Our research came up with numerous limitations. First, this study intended to build and evaluate prediction models by sorting patients at risk of FNI, not to construct a casual model. Hence, the rank of each variable does not necessarily display the importance of that variable. Despite the fact that data result from a multicentric collection, the retrospective nature of the collection along with the difficulty to evaluate surgeon skills, plus the relatively low event rate, added to the exclusion of revision surgery, parotid accessory lobe tumors, and malignant histology, represent a risk of inherent bias within our data set. Additionally, all possible residual measured or unmeasured confounders that could have influenced the outcomes should have been considered. However, it is necessary to underline that this research means to be the first multicentric study concerning FNI after PGS for benign tumors, developed to measure and evaluate the effectiveness and performance of these algorithms in diverse surgical environments.

An ongoing process, which includes a web-based application grounded on our algorithm performance is under construction. Such application will enable us to test our model's capability to design predictions in an actual clinical setting through prospective data collection and therefore, enhance its forecasting ability by including anonymized clinical data from head and neck clinical centers around the world.

CONCLUSION

This research validates that ML models like RF and ANN may deliver evidence-based forecasting from multicentric data concerning the risk of FNI. With the advent of ML technology, the implementation of clinical, radiological, histological, and/or cytological data may result in an improvement and amplitude of the information regarding the potential risks of FNI associated with patients before each surgery. Interdependent on the current results over multicentric data, prospective studies that advocate for the evaluation and analysis of the actual possibilities of this procedure are necessary to corroborate and ensure the implementation of this technology in a tangible clinical scenario.

ACKNOWLEDGMENTS

None.

AUTHOR CONTRIBUTIONS

Carlos M. Chiesa-Estomba, Jose A. González-García, Ekhiñe Larruscain, Jon A. Sistiaga Suarez, Paula Martínez-Ruiz de Apodaca, and Alfonso Medela substantially contributed to the conception or design of the work; acquisition, analysis, or interpretation of data for the work. Carlos M. Chiesa-Estomba, Xavier León, and Alfonso Medela drafted or revised the work critically for important intellectual content. Carlos M. Chiesa-Estomba, Miquel Quer, Xavier León, Paula Martínez-Ruiz de Apodaca, Miguel Mayo-Yanez, and Alfonso Medela approved the final version to be published. Celia López-Mollá agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Not applicable.

ETHICS STATEMENT

Approval from the ethics committee was obtained (CCH-110419).

ORCID

Carlos M. Chiesa-Estomba http://orcid.org/0000-0001-9454-9464

REFERENCES

1. Nitzan D, Kronenberg J, Horowitz Z, et al. Quality of life following parotidectomy for malignant and benign disease. *Plast Reconstr Surg*. 2004;114:1060-1067.
2. Ryzenman JM, Pensak ML, Tew JM, Jr. Facial paralysis and surgical rehabilitation: a quality of life analysis in a cohort of 1,595 patients after acoustic neuroma surgery. *Otol Neurotol*. 2005;26:516-521, discussion 521.
3. Bur AM, Shew M, New J. Artificial intelligence for the otolaryngologist: a state of the art review. *Otolaryngol Head Neck Surg*. 2019;160:603-611.
4. Zheng M-H, Shi K-Q, Lin X-F, et al. A model to predict 3-month mortality risk of acute-on-chronic hepatitis B liver failure using artificial neural network. *J Viral Hepatitis*. 2013;20:248-255.
5. Biglarian A, Hajizadeh E, Kazemnejad A, Zayeri F. Determining of prognostic factors in gastric cancer patients using artificial neural networks. *Asian Pacific J Cancer Prev*. 2010;11:533-536.
6. Biglarian A, Hajizadeh E, Kazemnejad A, Zali M. Application of artificial neural network in predicting the survival rate of gastric cancer patients. *Iran J Publ Health*. 2011;40:80-86.
7. Amiri Z, Mohammad K, Mahmoudi M, Zeraati H, Fotouhi A. Assessment of gastric cancer survival: using an artificial hierarchical neural network. *Pak J Biol Sci*. 2008;11:1076-1084.
8. Faradmal J, Soltanian AR, Roshanaei G, Khodabakhshi R, Kasaeian A. Comparison of the performance of log-logistic regression and artificial neural networks for predicting breast cancer relapse. *Asian Pacific J Cancer Prev*. 2014;15:5883-5888.
9. Abouzari M, Goshtasbi K, Sarna B, et al. Prediction of vestibular schwannoma recurrence using artificial neural network. *Laryngoscope Investig Otolaryngol*. 2020;5:278-285.
10. Alabi RO, Elmusrati M, Sawazaki-Calone I, et al. Machine learning application for prediction of locoregional recurrences in early oral tongue cancer: a web-based prognostic tool. *Virchows Arch*. 2019;475:489-497.
11. Chiesa-Estomba CM, Echaniz O, Sistiaga Suarez JA, et al. Machine learning models for predicting facial nerve palsy in parotid gland surgery for benign tumors. *J Surg Res*. 2021;262:57-64.
12. Chiesa-Estomba CM, Sistiaga-Suarez JA, González-García JÁ, et al. Artificial neural network as a tool to predict facial nerve palsy in parotid gland surgery for benign tumors. *Med Sci*. 2020;8:42.
13. Quer M, Guntinas-Lichius O, Marchal F, et al. Classification of parotidectomies: a proposal of the European Salivary Gland Society. *Eur Arch Otrhinolaryngol*. 2016;273:3307-3312.
14. Quer M, Vander Poorten V, Takes RP, et al. Surgical options in benign parotid tumors: a proposal for classification. *Eur Arch Otrhinolaryngol*. 2017;274:3825-3836.
15. Guntinas-Lichius O, Gabriel B, Peter Klussmann J. Risk of facial palsy and severe Frey's syndrome after conservative parotidectomy for benign disease: analysis of 610 operations. *Acta Otolaryngol*. 2006;126:1104-1109.
16. Witt RL. Facial nerve monitoring in parotid surgery: the standard of care. *Otolaryngol Head Neck Surg*. 1998;119:468-470.
17. Dulguerov P, Marchal F, Lehmann W. Postparotidectomy facial nerve paralysis: possible etiologic factors and results with routine facial nerve monitoring. *Laryngoscope*. 1999;109:754-762.
18. Mehle ME, Kraus DH, Wood BG, et al. Facial nerve morbidity following parotid surgery for benign disease: the Cleveland Clinic Foundation experience. *Laryngoscope*. 1993;103:386-388.
19. Bron LP, O'Brien CJ. Facial nerve function after parotidectomy. *Archiv Otolaryngol Head Neck Surg*. 1997;123:1091-1096.
20. Mra Z, Komisar A, Blaugrund SM. Functional facial nerve weakness after surgery for benign parotid tumors: a multivariate statistical analysis. *Head Neck*. 1993;15:147-152.
21. Ward CM. Injury of the facial nerve during surgery of the parotid gland. *Br J Surg*. 1975;62:401-403.
22. Yuan X, Gao Z, Jiang H, et al. Predictors of facial palsy after surgery for benign parotid disease: multivariate analysis of 626 operations. *Head Neck*. 2009;31:1588-1592.
23. Ikoma R, Ishitoya J, Sakuma Y, et al. Temporary facial nerve dysfunction after parotidectomy correlates with tumor location. *Auris Nasus Larynx*. 2014;41:479-484.
24. Domenick NA, Johnson JT. Parotid tumor size predicts proximity to the facial nerve. *Laryngoscope*. 2011;121:2366-2370.
25. Grosheva M, Pick C, Granitzka T, et al. Impact of extent of parotidectomy on early and long-term complications: a prospective multicenter cohort trial. *Head Neck*. 2019;41:1943-1951.
26. Jin H, Kim BY, Kim H, et al. Incidence of postoperative facial weakness in parotid tumor surgery: a tumor subsite analysis of 794 parotidectomies. *BMC Surg*. 2019;19:199.
27. Karadaghy OA, Shew M, New J, Bur AM. Development and assessment of a machine learning model to help predict survival among patients with oral squamous cell carcinoma. *JAMA Otolaryngol Head Neck Surg*. 2019;145:1115-1120.
28. Yang CQ, Gardiner L, Wang H, Hueman MT, Chen D. Creating prognostic systems for well-differentiated thyroid cancer using machine learning. *Front Endocrinol*. 2019;10:288.

29. Shew M, New J, Bur AM. Machine learning to predict delays in adjuvant radiation following surgery for head and neck cancer. *Otolaryngol Head Neck Surg*. 2019;160:1058-1064.
30. Fujima N, Shimizu Y, Yoshida D, et al. Machine-learning-based prediction of treatment outcomes using MR imaging-derived quantitative tumor information in patients with sinonasal squamous cell carcinomas: a preliminary study. *Cancers*. 2019;11:800.
31. Bur AM, Holcomb A, Goodwin S, et al. Machine learning to predict occult nodal metastasis in early oral squamous cell carcinoma. *Oral Oncol*. 2019;92:20-25.
32. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media; 2009.
33. Schmueli G. To explain or to predict? *Stat Sci*. 2010;25:289-310.
34. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods*. 2018;15:233-234.
35. Schroeder LD, Sjoquist DL, Stephan PE. *Understanding Regression Analysis*. Sage; 1986.

How to cite this article: Chiesa-Estomba CM, González-García JA, Larruscain E, et al. Facial nerve palsy following parotid gland surgery: a machine learning prediction outcome approach. *World J Otorhinolaryngol Head Neck Surg*. 2023;9:271-279.
doi:10.1002/wjo2.94