



OPEN

Decision making on vestibular schwannoma treatment: predictions based on machine-learning analysis

Oliver Profant², Zbyněk Bureš⁵✉, Zuzana Balogová³, Jan Betka⁴, Zdeněk Fík⁴, Martin Chovanec³ & Jan Voráček¹

Decision making on the treatment of vestibular schwannoma (VS) is mainly based on the symptoms, tumor size, patient's preference, and experience of the medical team. Here we provide objective tools to support the decision process by answering two questions: can a single checkup predict the need of active treatment?, and which attributes of VS development are important in decision making on active treatment? Using a machine-learning analysis of medical records of 93 patients, the objectives were addressed using two classification tasks: a time-independent case-based reasoning (CBR), where each medical record was treated as independent, and a personalized dynamic analysis (PDA), during which we analyzed the individual development of each patient's state in time. Using the CBR method we found that Koos classification of tumor size, speech reception threshold, and pure tone audiometry, collectively predict the need for active treatment with approximately 90% accuracy; in the PDA task, only the increase of Koos classification and VS size were sufficient. Our results indicate that VS treatment may be reliably predicted using only a small set of basic parameters, even without the knowledge of individual development, which may help to simplify VS treatment strategies, reduce the number of examinations, and increase cause effectiveness.

Vestibular schwannoma (VS) is the most common tumor of the temporal bone. It is a benign, mostly solitary and slowly growing tumor that grows from the Schwann cells of the vestibular portion of the 8th cranial nerve. VS causes approximately 80% of the tumors of the pontocerebellar angle, and around 8–10% of intracranial tumors¹. The symptomatology of VS is mainly caused by the compression or destruction of the surrounding structures, and an obstruction in the flow of cerebrospinal fluid, and comprise mainly asymmetric hearing loss^{2,3}, unilateral tinnitus⁴, or balance disorders and cepalea⁵.

Basically, there are two possible approaches to a patient with a VS: a wait-and-scan (WaS) strategy during which the patient undergoes regular checkups with no active treatment, and an active treatment of the tumor. A long WaS monitoring might eventually lead to an increased tumor size and subsequent complicated operation; however if there is no VS progress, such conservative treatment is economic and harmless to the patient. The active treatment (surgery or radiotherapy) is more beneficial in smaller tumors⁶. Although there is always a chance that the tumor will not grow and no intervention would be necessary, the length of postponement of active intervention (even with relatively small tumor growth) can worsen the results^{7–9}. Therefore, an untimely decision on active treatment might lead to poorer results and unnecessary costs.

At the initial diagnosis and during the subsequent regular checkups, a number of diagnostic variables is gathered. Based on these variables and their dynamics, a decision on further treatment is made. However, contributions of the individual variables to the final decision may vary; furthermore, for some variables the static values are important, while for other variables the dynamic change is the key. Knowledge of these principles is

¹Faculty of Management, Prague University of Economics and Business, Jindřichuv Hradec, Czech Republic. ²Department of Auditory Neuroscience, Institute of Experimental Medicine, Czech Academy of Sciences, Prague, Czech Republic. ³Department of Otorhinolaryngology, 3rd Faculty of Medicine, University Hospital Královské Vinohrady, Charles University in Prague, Prague, Czech Republic. ⁴Department of Otorhinolaryngology and Head and Neck Surgery, 1st Faculty of Medicine, University Hospital Motol, Charles University in Prague, Prague, Czech Republic. ⁵Department of Cognitive Systems and Neurosciences, Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University, Jugoslávských partyzánů 1580/3, 160 00 Prague 6, Czech Republic. ✉email: zbynek.bures@cvut.cz

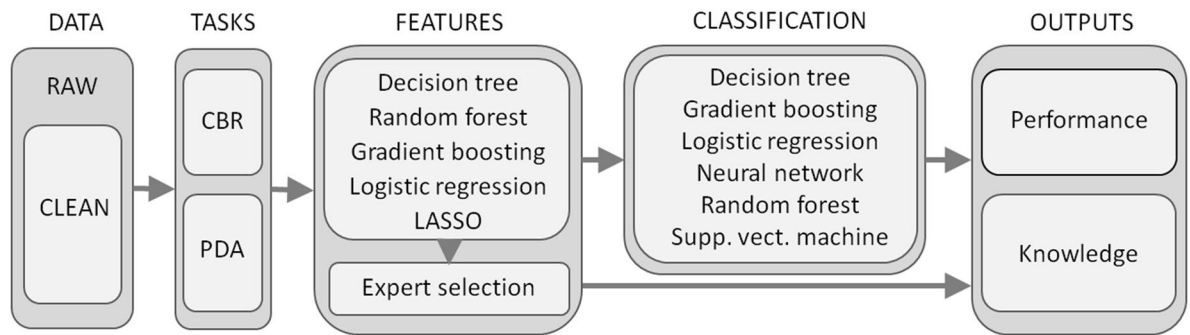


Figure 1. An outsketch of the methodological process used in the analysis. After cleaning the data, the problem was solved in two parallel tasks (CBR and PDA). Using several feature selection methods followed by expert evaluation, the most important predictors of active VS treatment were identified. The identified set of predictors was processed by several classification methods to create models capable of predicting the active VS treatment based on the predictor values. The performance of the models was analyzed using various performance metrics.

important in two ways: it could optimize the diagnostic routine by eliminating the unnecessary procedures, and it could support the medical teams in their decisions by providing an objective reasoning of the patient's state.

Machine learning techniques represent a promising tool for supporting decisions in many disciplines. Statistical processing seeks quantitative identification and an explanation of relationships among variables, however, the precision and reliability of the statistical description strongly depends on a priori assumptions and the size of the data sample. This is particularly limiting when it comes to multidimensional data. The approach of artificial intelligence (to which the machine learning belongs) can overcome these limitations by building a model using known training data, which is subsequently validated using validation data. This model is then utilizable for making predictions or decisions; its performance (correctness of its decisions or predictions) can be further assessed by testing data that were not available to the model during the learning phase.

The aim of this study was to address the following questions: (1) can a single examination (for example, the initial checkup at the time of VS diagnosis) reliably predict the need for active treatment? (2) If so, what are the diagnostic variables and their values that can lead to such a prediction? (3) When evaluating the dynamics of the patient's state, which changes of which variables are the most important ones for the decision on further treatment? We address these issues using machine-learning methods of data classification¹⁰, which is a promising analytical tool particularly in situations when the classical statistical processing is not suitable, e.g., due to extensive data dimensionality, insufficient size of the data sample, or when the necessary a priori assumptions are not met. We approach the problem from two viewpoints. First, we treat each checkup record as an independent entity and analyze which checkups resulted in a decision of active treatment (the so called case-based reasoning, CBR). Second, we take into account the dynamic changes of all the diagnostic variables of each patient and look for those dynamic changes that best characterize the actively treated patients (the so called personalized dynamic analysis, PDA). Data sets for both problems were processed with supervised machine learning methods to identify and justify the most reliable predictors of VS treatment. In both tasks, we seek the minimum set of variables (features) along with their values (static or dynamic), that lead to the most reliable prediction of active treatment. As a result, we present for each task a black-box automated classifier that predicts the active treatment when provided with the appropriate data, and also a transparent set of rules based on a decision tree. An overview of the methodology is shown in Fig. 1. It is important to note that our conclusions were derived entirely from a strictly cleaned data set, which contained no subjective or methodological assumptions that could possibly affect the discovered information. Such unbiased resulting structures can serve as a ground truth, either for subsequent expert evaluations or for the comparison of results with more knowledge-intensive approaches, including statistics.

Methods

We present the results of a semi-supervised analysis of 388 medical records, characterizing the wait-and-scan (WaS) phase of vestibular schwannoma development for 93 individually followed VS patients. Our group of patients was selected from approximately 400 patients with diagnosed VS, examined at the Department of Otorhinolaryngology and Head and Neck Surgery, 1st Faculty of Medicine, Charles University, University Hospital Motol between 2012 and 2018. The main input criterion was the selection for the WaS protocol based on the initial examination. The original set of source diagnostic variables was cleaned and restructured.

Data acquisition. Diagnostic data were obtained at the Department of Otorhinolaryngology and Head and Neck Surgery, 1st Faculty of Medicine, Charles University, University Hospital Motol, between 2012 and 2018. The examination procedures, and the informed consent, were approved by the Ethics Committee of the University Hospital Motol, in Prague. All the participants provided their written informed consent to participate in this study; signed written consents are stored at the Department. All procedures were performed in accordance with relevant guidelines and regulations and with the Declaration of Helsinki.

Data characteristics. The original data set included 388 records of 93 patients (55 females, 38 males; age median 59 years, 44 left-sided VS, 49 right-sided VS). For the 53 patients who were retained in the wait-and-scan regime, the median duration of the overall investigation period was 51 months inclusive of 5 checkups. Within the actively treated group, the median duration of the wait-and-scan regime lasted 37 months and required 3 checkups.

The raw data obtained by commonly used diagnostic techniques were organized in a table, where each row represented a single diagnostic checkup which either resulted in active treatment or not, and where columns corresponded to diagnostic variables as follows:

- Pure tone audiometry [PTA (dB)]—pure-tone hearing thresholds measured separately for each ear at eight frequencies from 0.25 to 8 kHz in attenuated chamber,
- Speech audiometry (measured in the diseased ear in attenuated chamber)—Speech reception threshold [SRT (dB)], Speech discrimination score [SDS (%)], Maximum discrimination level [MDL (dB)], and Maximum discrimination ratio [MDR (%)],
- Magnetic resonance imaging-based descriptors (the size was evaluated on T2-weighted MRI):
 - Size of VS: maximal 1D size [mm],
 - Koos grading (Class 1–4).
- Derived row-based metrics for CBR, calculated from PTAs separately for each ear, and separately for two frequency ranges (full—whole set of frequencies up to 8 kHz, and basic—only frequencies up to 4 kHz):
 - average PTA in dB, denoted as PTA_{XARn} where X can either be VS (diseased ear) or H (healthy ear), and n is either 8 (full range) or 4 (basic range); for example, PTA_{VSAR8} is the average PTA of the diseased ear computed from frequencies up to 8 kHz,
 - slope and intercept of linear fit of pure-tone thresholds in dB, denoted as PTA_{XSRn} and PTA_{XIRn} , respectively
 - difference of average PTA between the two ears, denoted as PTA_{DAR4} or PTA_{DAR8} .
 - The resultant data set had 184 38-dimensional records.
- Derived column-based metrics for PDA, calculated from time-dependent changes of selected variables (including the row-based ones) separately for each patient:
 - average, denoted as var_AC , where var is the variable from which the column-based average is computed,
 - slope, denoted as var_SC , for example PTA_{DAR4_SC} stands for time-dependent slope of the inter-ear difference of average PTA computed over the basic frequency range,
 - intercept, denoted as var_IC ,
 - last and total differences, denoted as var_LC and var_TD , respectively.
 - The resultant data set had 42 24-dimensional records.

Several other functions were examined in the patients (auditory brainstem response (ABR), otoacoustic emissions (OAE), vestibular function), however, they were either not recordable (ABR, OAE) or were not consistently provided over the course of time, therefore they were excluded from the current analysis.

Subjective characteristics of the patients, such as vertigo or tinnitus, were also gathered but were not included in the current analyses. The current study was designed as entirely non-parametric and data-driven; therefore to avoid any possible subjectivity we purposely suppressed the influence of non-deterministic factors, including the patients' subjective characteristics. For the same reason, all incomplete records were removed instead of artificially imputing the missing values. Additionally the phase of data transformation was omitted, as it usually leads to the normalization or equalization of data distributions. Although our restrictions caused the loss of some information, this approach avoids unjustified biases, is fully repeatable and extendable, and as such represents a core baseline model, which can later serve as a reliable benchmarking etalon for comparison with alternative ways of processing; namely including traditional parametric statistical techniques.

Data processing—general. The applied methodology follows the general Knowledge Discovery in Databases process, introduced in¹¹ or¹². The data were processed with supervised, internally transparent machine learning methods as follows:

- No a priori assumptions concerning the cumulative characteristics of data were made, so the presented results are not biased by any artificial modifications, like imputations or transformations.
- Only complete records were selected for further processing.
- Two complementary approaches: (1) static, anonymized CBR, and (2) personalized PDA, were applied to discover knowledge hidden in a multi-dimensional space.
- CBR assigns single medical records (rows) to the binary target decisions on the treatment (WaS/active), it considers neither the characteristics of individual patients, nor their history of VS progress.

- PDA also performs binary classification, but works with the temporal courses of selected variables taken from the complete WaS checkup history of single patients. Thus, every processed sample summarizes the complete column-wise WaS history of the given patient.
- An interactive reduction of dimensionality (feature selection) preserving the meaning and relations among the original variables was performed, to exclude the less significant features, and simplify the problem and increase the generalization capabilities of the resulting structures.
- The data for all supervised learning tasks were equally balanced with respect to the target, and randomly divided into the training, validation, and test sets in the proportion 50:30:20. The first two partitions were used for learning and optimization of the desired type of discrimination function, the last subset contained unseen data and served for the numeric evaluation of classification performance.

The supervised elimination of redundant features. An initial reduction of dimensionality was performed in all classification tasks, using the five below-listed techniques implemented with the StatExplore, HP Random Forest, Gradient Boosting, Variable Selection and HP Variable Selection nodes of SAS Enterprise Miner:

- (1) Decision or classification tree^{13–15} with Chi-square split¹⁶.
- (2) Random forest^{17,18} with the Gini impurity index G ^{19,20} as the node splitting metrics.
- (3) Gradient boosting^{21,22} using Gini impurity index for updating the decision tree.
- (4) Logistic regression^{23–25} with respect to the target class, applied to the results of forward stepwise regression²⁶ of a gradually reduced set of pairwise (R-squared) correlations²⁷.
- (5) Least absolute shrinkage and selection operator (LASSO)^{28,29}.

In addition to these algorithms, an expert (manual) selection of the most significant features was performed, which is also the main output from the knowledge elicitation phase. At the end of this iterative process, we proposed the minimal set of variables efficiently characterizing the analyzed problem, based on the outputs of the previous five algorithmic methods. The primary criterion for selection of a given variable was its occurrence among the best ten candidates, which must be either greater or equal to 3, or its average ranking lower or equal to 5. Perspective combinations of such preliminarily selected candidates were interactively analyzed, to eliminate the least significant members and maximize the credibility of the discovered knowledge.

Supervised learning and classification. In the classification stage we used the following techniques:

- (1) Decision tree, random forest, gradient boosting and logistic regression, all referred to in the previous section.
- (2) Support vector machine with radial basis function kernel^{30–32}.
- (3) Feed-forward neural network^{33,34}.

The optimal classifier was selected as the best performing combination of the six feature selection techniques given in the previous section (logistic regression, decision tree, random forest, gradient boosting, LASSO, and interactive expert selection) with the six types of classifiers given here.

Performance metrics. To evaluate classification performance, several indicators were used:

- Accuracy (ACC)—the rate of correct classification for the evaluated data set:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{P + N}$$

where TP is the true positive, TN is the true negative, FP is the false positive, FN is the false negative, P is the all real positive ($P = TP + FN$), N is the all real negative ($N = TN + FP$) cases.

- Sensitivity (also recall or true positive rate, TPR)—the ability to correctly classify TP cases:

$$TPR = \frac{TP}{TP + FN} = \frac{TP}{P}$$

- Specificity (also selectivity or true negative rate, TNR)—the ability to correctly classify TN cases:

$$TNR = \frac{TN}{TN + FP} = \frac{TN}{N}$$

- Precision (also positive predictive value, PPV)—the rate that the predicted positive is TP:

$$PPV = \frac{TP}{TP + FP}$$

- Area under the Receiver operating characteristic curve (AUC)^{35,36}. Practically applicable classifiers should have $AUC > 0.6$, while $AUC > 0.9$ indicates an excellent performance.

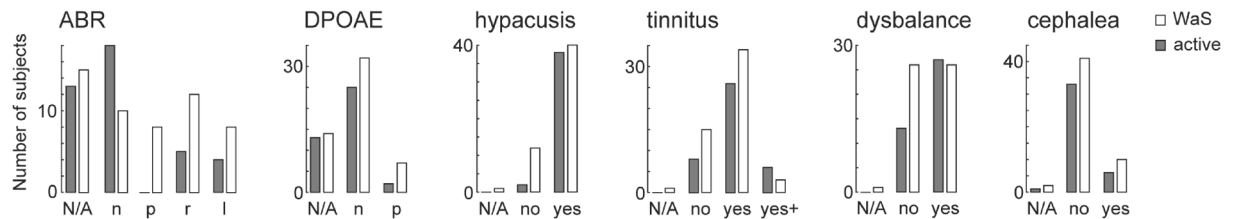


Figure 2. Diagnostic data of the patients included in the analysis. The bars represent the number of subjects having a certain characteristic. *N/A* not available; *n* ABR/DPOAE response not present; *p* ABR/DPOAE response present; *r* ABR with signs of retrocochlear lesion; *l* ABR with prolonged latencies. *Yes +* annoying tinnitus. Gray bars—actively treated patients; white bars—wait-and-scan patients.

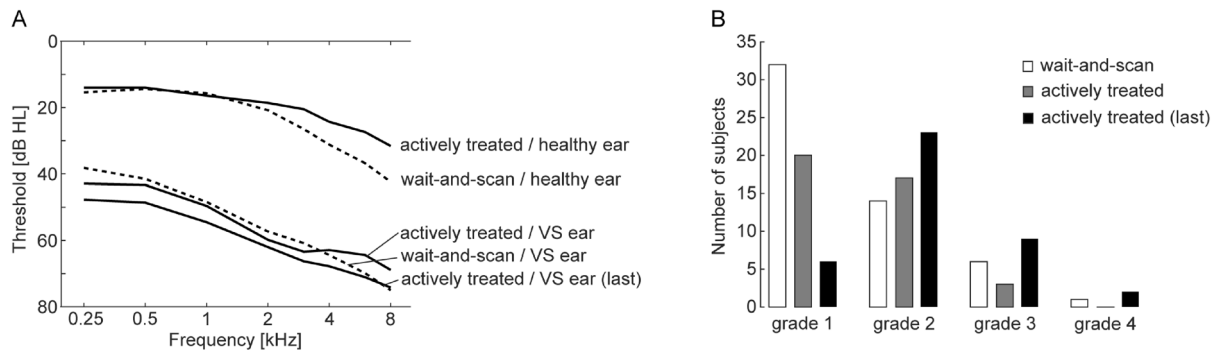


Figure 3. Hearing thresholds and tumor sizes of the patients included in the analysis. **(A)** Average audiograms recorded in the healthy and diseased ears in wait-and-scan patients and in the patients later changed to active treatment during the initial examination, plus the average audiogram of the diseased ear in the actively treated subjects recorded immediately before the change from wait-and-scan to active treatment. **(B)** Histogram of Koos grades identified in the actively treated and wait-and-scan patients during the initial examination, and in the actively treated subjects recorded immediately before the change from wait-and-scan to active treatment, the bars represent numbers of subjects having a certain Koos grade.

- Average square error (ASE)—squared metric difference between the target and continuous output of the discrimination function, divided by the number of samples.

Results

The general diagnostic data of the patients included in the analysis are illustrated in Fig. 2. These graphs show the number of subjects having a certain result of ABR and distortion products of OAE (DPOAE) examinations, as well as subjective characteristics such as hypacusis or tinnitus. Figure 3A depicts averaged audiograms recorded from both healthy and VS ears during the initial examination, plus the average audiogram of the diseased ears recorded immediately before the change from wait-and-scan to active treatment. Figure 3B shows the histogram of Koos grades recorded during the initial examination in wait-and-scan patients, patients who were later changed to active treatment, and in the actively treated patients recorded immediately before the change from wait-and-scan to active treatment.

The section below summarizes the results of the two interrelated analytic phases, dimensionality reduction including knowledge extraction, and supervised learning for both CBR and PDA experiments.

CBR—dimensionality reduction and knowledge extraction. The output of this method is a set of the most important diagnostic characteristics (variables) along with their significant values. The method aims to provide a transparent set of rules which, using the values of the selected variables, can simply be used generally to support the decision on VS treatment.

Initially, the dimensionality of the full set of CBR variables was reduced with five algorithmic methods (see Table 1). Each of the methods provided 10 variables, rated as the most important for the prediction of VS treatment. Using the variables suggested by the algorithmic methods we manually performed an expert ranking, resulting in an initial version of a reduced set of variables (denoted as CBR_{EXPINI}). By interactive minimizations of this initial set we finally proposed a minimum set of variables (CBR_{EXPFIN}), necessary for the reliable prediction of VS treatment. Table 2 shows the performance for different sets of variables; it is obvious that the removal of unnecessary variables actually improves the prediction accuracy, and furthermore, the output generated by the expertly found features is comparable with the average performance of the three best automated supervised classifiers and feature selectors marked as CBR_{CLASS} (see Tables 5 and 6 in the next section). In addition, Table 2 presents the quality of adaptation on known samples (an average of performance on training and validation data).

Ord	Decision tree	Random forest	Gradient boosting	Logistic regression	LASSO	Expert selection (CBR _{EXP})			
						Initial	Num	Avg	Final
1	PTA _{VS} SR8	Koos	Koos	Koos	Koos	Koos	5	1.2	Koos
2	Koos	PTA _{VS} SR8	PTA _{VS} SR8	Size	PTA _H SR8	SRT	5	3.8	SRT
3	PTA _D AR4	SRT	PTA _H SR8	SRT	Size	PTA _{VS} SR8	4	3.3	PTA _{VS} SR8
4	SRT	PTA _{VS} AR4	SRT	PTA _{VS} 0.25	PTA _{VS} SR4	PTA _H SR8	4	4.0	PTA _H SR8
5	PTA _H SR8	PTA _{VS} 3	Size	PTA _H 8	SRT	Size	4	5.0	-
6	PTA _{VS} SR4	PTA _H SR8	PTA _{VS} IR8	PTA _H 0.5	PTA _H IR8	PTA _{VS} SR4	3	5.7	-
7	PTA _D AR8	PTA _{VS} SR4	PTA _D AR4	PTA _H 2	SDS	PTA _{VS} 0.25	3	7.0	-
8	PTA _H IR8	PTA _{VS} 0.25	PTA _D AR8	PTA _{VS} 6	PTA _{VS} SR8	PTA _H IR8	3	7.7	-
9	PTA _{VS} 3	MDL	PTA _H IR8	PTA _{VS} 8	PTA _{VS} 0.25	PTA _D AR4	2	5.0	PTA _D AR4
10	Size	PTA _H 8	PTA _{VS} 1	SDS	MDR	-	-	-	-

Table 1. Predictors, extracted from CBR data, ordered according to their significance for applied dimensionality reduction method. Irrelevant variables were rejected from expert selection. Variables and metrics are expressed as follows (see “Methods”): (a) suffix 4 holds for the basic frequency range, suffix 8 holds for the full frequency range; (b) subscripts H and VS hold for the healthy or diseased ear, respectively, subscript D holds for the difference of averaged PTA values between the two ears; (c) tailing abbreviations have the following meaning: *AR* average row wise, *IR* intercept row wise, *SR* slope row wise.

Variables		Train and validation (AVG, %)						Test (%)					
Set	Number	ACC	PPV	TPR	TNR	AUC	ASE	ACC	PPV	TPR	TNR	AUC	ASE
CBR	38	78	75	84	72	87	15	72	69	79	65	81	19
CBR _{EXP} INI	9	82	79	88	76	84	15	76	72	83	69	76	19
CBR _{EXP} FIN	5	84	81	90	78	88	14	81	78	88	74	78	17
CBR _{CLASS}	38	86	84	91	82	92	11	88	85	93	82	87	12

Table 2. Performance of gradually reduced expert set of variables for CBR data.

Based on the aforementioned findings, we can claim that knowledge of the Koos classification, SRT, and three PTA-derived variables, provides sufficient information for a reliable VS surgery decision; even in the case of a single medical checkup. Therefore it may be feasible to exclude clinical tests of the less significant features, which can make the daily diagnostic routine faster and cheaper.

Using the individual variable values, it is now possible to decide whether to perform active treatment (*Yes* decision) or not (*No* decision). An important question is what the boundary values of the variables are, i.e., at which level each variable switches the decision from No to Yes. The answer, however, is not unique because the selected features can be assigned into numerous structurally different solutions with comparable performances. One possible solution is given in Fig. 4, and in detail in Table 3. By traversing this binary decision tree according to the rules, we finally arrive at the decision in the leaves; decision accuracy in the leaf nodes is approximately 80%. The ability of the decision tree to also handle missing (N/A) values is yet another advantage of this technique. An example of several CBR records taken from our data and the corresponding decisions is shown in Table 4.

These experimental results confirmed the applicability of the variable set CBR_{EXP}FIN for the reliable predictions of VS surgery. The presented structural representation (i.e., the decision tree in Fig. 4) can help practitioners in a more informed analysis of diagnostic results.

CBR—supervised learning. The previous method gave a transparent set of significant variables and their values that can be directly used for the prediction or decision on VS treatment. However, its result is generally ambiguous; furthermore, our intention to minimize the variable set as much as possible might lead to a certain loss of accuracy. For such reasons, we also decided to create a black-box-like solution based on an automated feature selector followed by a classifier. We identified and parametrized perspective combinations of the six feature selectors with six classifiers. As in the previous method, the CBR data set was split into the training, validation, and test partitions, and batch processed for all the 36 combinations of feature selectors and classifiers. The results of classification accuracy are summarized in Table 5.

Table 5 shows that the gradient boosting algorithm is on average the best performing algorithm for both data processing phases (i.e., it works the best both as a feature selector and as a classifier). The globally best result was generated by its combination with neural network (89%). The performance of the fixed expert selection of variables in the CBR_{EXP}FIN set is also remarkable, particularly when followed by a gradient boosting classifier.

Full results of the three best performing combinations are shown in Table 6, the corresponding Receiver operating characteristic (ROC) curves are depicted in Fig. 5. The slightly worse performance for the train and

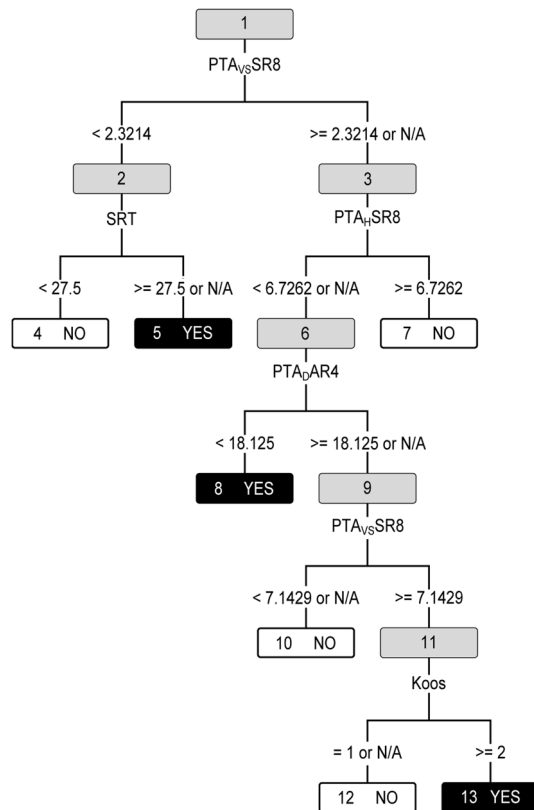


Figure 4. CBR decision tree. A decision tree for CBR_{EXP}FIN variables, applied on CBR data.

Node ID	Samples	Node rule	No	Yes
1	184	$PTA_{V3}SR8 \geq 2,3$ or N/A	2	3
2	61	$SRT \geq 27.5$ or N/A	4	5
3	123	$PTA_{H1}SR8 \geq 6.7$	6	7
4	9	No		
5	52	Yes		
6	86	$PTA_{D}AR4 \geq 18.1$ or N/A	8	9
7	37	No		
8	17	Yes		
9	69	$PTA_{V3}SR8 \geq 7.1$	10	11
10	39	No		
11	30	Koos = 2	12	13
12	18	No		
13	12	Yes		

Table 3. Tabular representation of a decision tree for CBR_{EXP}FIN variables, applied on CBR data.

validation set, in comparison with the test set, was caused by a larger validation error. However, as the key performance indicator was behavior for unknown test data, we accepted this local decrease which was mainly caused by a small number of learning samples in comparison with the number of significant variables. Regardless, Table 6 shows that the absolute test accuracies, as well as biases and variances of the winning combinations, are sufficient for daily use.

To compare the results obtained from the traditional, two-stage processing with those obtained from a complementary one-shot algorithm, we processed the full CBR dataset with the Deep learning algorithm. A set of experiments employing this modern technique was performed on a fully connected three-layered network. The layers included 38, 76, and 2 neurons with the rectified linear activation function. The network was trained with gradient descent back-propagation method. Such paradigm resulted in the following best performance:

ID	Koos	SRT	PTA _{vs} SR8	PTA _H SR8	PTA _D AR4	Real target	Nodes visited	Prediction
1	1	57	11.3	1.3	28.8	No	1, 3, 6, 9, 11, 12	No
2	1	30	4.7	0.2	5.0	Yes	1, 3, 6, 8	Yes
3	2	110	10.7	7.6	43.8	No	1, 3, 7	No
4	2	37	2.1	0.1	16.3	Yes	1, 2, 5	Yes
5	3	60	4.6	9.8	17.5	No	1, 3, 7	No
6	3	110	2.1	3.3	26.3	Yes	1, 2, 5	Yes
7	4	110	5.2	2.3	68.8	No	1, 3, 6, 9, 10	No
8	4	110	0	3.0	98.8	Yes	1, 2, 5	Yes

Table 4. Inferences for selected sample CBR records using a decision tree learned from CBR_{EXP}FIN variables.

Feature selection method	Classification accuracy (ACC) for test set (%)						
	Decision tree	Rand. Forest	Gradient boosting	Logistic regression	Supp. vect. machine	Neural network	Avg
Decision tree (CBR)	76	79	84	82	79	76	79
Random forest (CBR)	76	79	87	76	76	68	77
Gradient boosting (CBR)	84	82	86	84	76	89	84
Logistic regression (CBR)	67	78	68	86	83	85	78
LASSO (CBR)	74	79	79	76	74	82	77
Expert (CBR _{EXP} FIN)	76	76	87	79	76	74	78
Avg	76	79	82	81	77	79	79

Table 5. Performance of applied combinations of dimensionality reduction and classification techniques on test set for CBR and CBR_{EXP}FIN data.

Classification (feature selection)	Train and validation (avg, %)						Test (%)					
	ACC	PPV	TPR	TNR	AUC	ASE	ACC	PPV	TPR	TNR	AUC	ASE
Neural network (Gradient boosting)	84	82	88	80	93	10	89	87	93	85	92	8
Gradient boosting (CBR _{EXP} FIN)	84	81	89	79	89	13	87	84	94	80	86	15
Gradient boosting (random forest)	91	89	96	86	94	10	87	84	93	81	84	14

Table 6. Detailed metrics for the three best performing classifiers for CBR data.

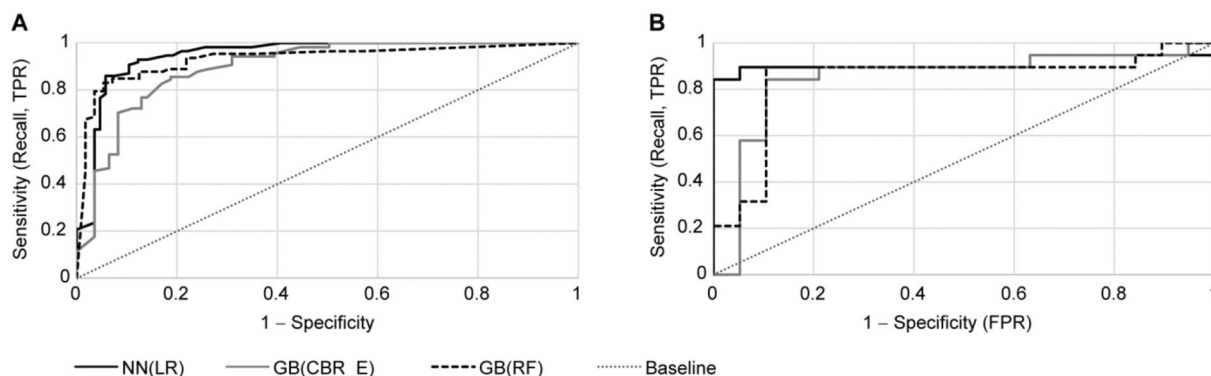


Figure 5. Receiver operating characteristic (ROC) curves of the three best performing classifiers for CBR data. (A) Averaged ROC curves of training and validation sets, (B) ROC curves for the test sets.

Ord	Decision tree	Random forest	Gradient boosting	Logistic regression	LASSO	Expert selection (PDA _{EXP})			
						Initial	Num	Avg	Final
1	Size_LD	Size_LD	Koos_LD	Size_LD	Koos_LD	Koos_LD	5	2	Koos_LD
2	Koos_LD	Koos_LD	Size_SC	PTA _{VS} AR4_SC	PTA _D AR4_IC	Size_LD	4	1.5	-
3	Size_AC	Size_SC	Size_LD	Size_SC	SRT_IC	Size_SC	4	4.5	Size_SC
4	PTA _{VS} AR4_AC	SRT_LD	PTA _{VS} SC4	Koos_LD	SRT_LD	Koos_TD	3	6.7	-
5	Koos_TD	Koos_TD	SRT_SC	PTA _D AR8_SC	PTA _{VS} AR8_IC	PTA _D AR4_SC	3	7.7	PTA _D AR4_SC
6	PTA _{VS} AR8_AC	PTA _{VS} AR8_IC	PTA _{VS} AR8_SC	PTA _{VS} AR8_SC	PTA _D AR8_LD	PTA _{VS} AR4_SC	3	5.3	PTA _{VS} AR4_SC
7	PTA _D AR8_IC	PTA _D AR4_SC	PTA _{VS} AR4_LD	PTA _{VS} AR8_SC	PTA _D AR8_IC	PTA _D AR8_IC	3	7.7	-
8	PTA _D AR4_SC	PTA _D AR4_LD	PTA _D AR8_SC	PTA _D AR4_SC	PTA _{VS} AR4_IC	Size_AC	3	7.3	-
9	SRT_AC	PTA _D AR8_AC	PTA _D AR8_LD	PTA _D AR8_IC	Size_AC	-	-	-	-
10	Size_SC	PTA _{VS} AR4_SC	Size_AC	Koos_TD	PTA _{VS} AR4_AC	-	-	-	-

Table 7. Predictors, extracted from PDA data set, ordered according to their significance for each dimensionality reduction method. Irrelevant variables were rejected from expert selection. Variables and metrics are expressed as follows (see “Methods”): (a) suffix 4 holds for the basic frequency range, suffix 8 holds for the full frequency range; (b) subscripts H and VS hold for the healthy or diseased ear, respectively, subscript D holds for the difference of averaged PTA values between the two ears; (c) tailing abbreviations have the following meaning: *AR* average row wise, *IR* intercept row wise, *SR* slope row wise, *AC* average column wise, *IC* intercept column wise, *SC* slope column wise, *LD* last difference, *TD* total difference.

Variables		Train and validation (Avg, %)						Test (%)					
Set	Number	ACC	PPV	TPR	TNR	AUC	ASE	ACC	PPV	TPR	TNR	AUC	ASE
PDA	24	87	84	92	82	88	12	82	79	89	75	81	16
PDA _{EXP} INI	8	87	84	92	82	88	12	82	79	89	75	81	16
PDA _{EXP} FIN	4	87	85	92	82	90	11	82	79	89	75	81	16
PDA _{CLASS}	24	85	82	91	79	83	14	90	88	95	86	100	10

Table 8. Performance of gradually reduced set of variables for the PDA data set.

$$ACC = 82\%, PPV = 78\%, TPR = 89\%, TNR = 75\%, AUC = 88\%, ASE = 13\%$$

which is slightly worse than performance of classifiers with separate feature selection and classification stages. This result was partially determined by low cardinality of the processed dataset, as the Deep learning approach is suitable particularly for processing of extensive multidimensional datasets.

PDA—dimensionality reduction and knowledge extraction. While the CBR data set and the corresponding methods generated their predictions based only on a single medical checkup, the PDA data set takes into account the individual history of checkups for each patient. It is evident that the time-dependent development of diagnostic variable values may bring important information into the decision process. Therefore, we also repeated the same ranking and specification procedures described for the CBR data set for the PDA data set, in order to minimize the number of input variables and to obtain a transparent set of decision rules. The variables suggested by the feature selectors and the structure of the resulting expert set (PDA_{EXP}) are shown in Table 7. Table 8 shows the detailed performance metrics of the gradually optimized variable set. As with the CBR data set, in this case we also see the positive effect of the lower number of inputs on the overall performance and primary role of size-oriented VS metrics.

The decision tree constructed from the PDA_{EXP}FIN variable set naturally suppressed both the PTA-related indicators, as is shown in Fig. 6 and Table 9. The result can be simply interpreted: if there is any change in Koos classification from the previous checkup, surgery is recommended. If the Koos class remains unchanged, the Size growth is checked and if the trend is positive, surgery is indicated. Generally, both identified variables are so significant that no other diagnostic procedures are necessary (neither the expertly identified PTA). Regardless, if they were performed, the results can enhance the existing CBR knowledge base.

PDA—supervised learning. Supervised PDA experiments suffered from the low number of samples and, consequently, the small size of the test set. Although this fact was efficiently compensated with the inherent dominance of both the size-related variables, test classification outputs were discretized into several levels, as obvious from Table 10. The overall weaker performance of the interactively selected set of features PDA_{EXP}FIN was caused by its fixed and relatively wide structure in comparison with the other dimensionality reduction tech-

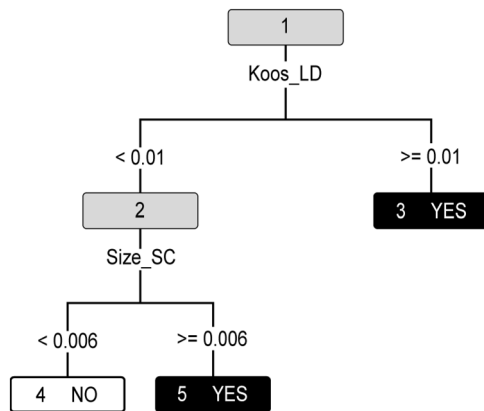


Figure 6. PDA decision tree. A decision tree for PDA_{EXP}FIN variables, applied on PDA data.

Node ID	Samples	Node rule	Yes	No
1	42	Koos_LD < 0.01	2	3
2	31	Size_SC < 0.006	4	5
3	11	Yes		
4	24	No		
5	7	Yes		

Table 9. Tabular representation of a decision tree for PDA_{EXP}FIN data set.

Feature selection method	Classification accuracy (ACC) for test set (%)						
	Decision tree	Rand forest	Gradient boosting	Logistic regression	Supp. vect. machine	Neural network	Avg
Decision tree (PDA)	80	80	80	90	80	70	80
Random forest (PDA)	80	70	80	70	90	70	77
Gradient boosting (PDA)	80	80	80	90	80	80	82
Logistic regression (PDA)	80	60	80	80	90	90	80
LASSO (PDA)	90	90	90	90	80	90	88
Expert (PDA _{EXP} FIN)	79	79	79	63	63	68	72
Avg	82	77	82	81	81	78	80

Table 10. Performance of applied combinations of dimensionality reduction and classification techniques on test set for PDA and PDA_{EXP}FIN data.

niques. In this specific situation, LASSO algorithm demonstrated the best average feature selection capabilities and its main component, logistic regression, as one of the most powerful classification algorithms on a global scale. Such conclusions correspond with the general knowledge concerning the classification of over-determined binary targets³⁷, and were also confirmed with the detailed characteristics of the best performing algorithms for the PDA task, presented in Table 11. Accordingly, the PDA data analysis confirmed the statement that the interim growth of VS itself, is the strongest and sufficient predictor of VS surgery.

Discussion

Over recent years, several studies have addressed the possibility of predicting VS growth, or a change from a conservative to an active treatment^{38–48}. Their outcomes are, however, ambiguous; some studies are inconclusive or fail to find any significant predictor of VS growth^{38,45}. The majority of the previous results state that the tumor size and also the degree of vestibular disorder are the key variables which influence the switch from conservative to active treatment. The above mentioned studies mostly analyzed the individual progress of symptoms, i.e., they worked in a manner similar to our PDA. Two studies specifically tested the hypothesis that VS growth could be predicted by the available data at diagnosis (i.e., the approach similar to our CBR); the study of Herwadker

Classification (feature selection)	Train and validation (avg, %)						Test (%)					
	ACC	PPV	TPR	TNR	AUC	ASE	ACC	PPV	TPR	TNR	AUC	ASE
Neural network (logistic regression)	85	82	91	79	88	14	90	88	95	86	100	10
Logistic regression (decision tree)	85	82	91	79	81	14	90	88	95	86	100	10
Logistic regression (gradient boosting)	85	82	91	79	80	15	90	88	95	86	100	10

Table 11. Detailed metrics for three best performing classifiers on PDA data set.

et al.⁴⁹ found no significant predictors, while Wolbers et al.⁵⁰ identified the long duration of hearing loss and intracranial localization of the tumor as the main predictors of a non-growing VS.

Here we present a novel approach to this issue which uses semi-supervised machine-learning techniques to create, parametrize, and evaluate four different models for the prediction of active treatment of vestibular schwannoma:

- (1) CBR—prediction from static variables
 - a. automated black-box classifier providing predictions given the input data
 - b. transparent set of rules (a decision tree) to support the decision on VS treatment
- (2) PDA—prediction from dynamic variables
 - a. automated black-box classifier providing predictions given the input data
 - b. transparent set of rules (a decision tree) to support the decision on VS treatment

The models were trained, validated, and tested using different subsets of the source data, which means that their performances (accuracy etc.) represent realistic values obtained with unknown data. In the applied methods, we concentrated on preservation of the original meaning of the individual attributes so that they remain transparent and interpretable during the entire classification process. This means that we used no multiplicative or other nonlinear transformations, but we instead employed only generalized linear models (LASSO, logistic regression, decision tree) and generalized (random) additive models, represented with the gradient boosting and random forest approaches. Although the latter two approaches are internally non-transparent, they still work with the original meaning of attributes.

The major findings state that using a simple decision tree it is possible to predict VS treatment, even from the static values of a few basic variables (Koos classification, speech reception threshold, and pure tone audiometry), with approximately 80% accuracy. Ultimately a higher accuracy (89%) can be achieved using a black-box classifier on the static data. From the dynamic point of view, we found that VS treatment can be predicted using dynamics of solely size-oriented variables (Koos classification and 1D size), both with a decision tree and with the black-box classifier. The prediction accuracy is slightly higher than that of the CBR approach.

Besides the provided prediction mechanisms alone, our analyses also indicate that only pure-tone hearing thresholds in both ears, speech reception threshold in the diseased ear, and Koos classification, are necessary at the first checkup (these variables are used in the static predictions); while during the subsequent follow-up, mainly the size-derived metrics and their dynamics play a role in the decision process. These findings might help to make the procedures related to the monitoring and treatment of VS patients more time- and cost-efficient, by eliminating the unnecessary measurements.

Supervised feature selection. The selection of the most important variables is essential in classification tasks where the number of available samples is comparable with the number of input variables, as over-fitted structures are characterized by the poor classification of unknown samples and low generalization ability^{51–54}. Considering that both CBR and PDA tasks belong to this category, an initial reduction of dimensionality was unavoidable. Employing the outputs of five dimensionality reduction techniques, we manually performed an expert selection of the most significant features. We believe that the final selection, numerically over-performing the initial configuration, optimally characterizes the key diagnostic symptoms, based on which the reliable VS surgery decision can be made at the very earliest.

Supervised learning and classification. The supervision in learning lies in the fact that the searched discrimination function is built from samples with a-priori known output membership. In contrast to the dimensionality reduction, internal interpretability of the learned classifier is not required, which results in a black-box-like nature. The previously introduced tree and regression-based techniques were re-used for the selection of significant variables but, as opposed to the manual interpretation of their results, performed in the feature selection process; this first stage was followed here by a learned classification algorithm.

The main mission of the classification task is the best performing inference, i.e. an accurate assignment of real-world clinical data to the predefined classes (in our case, wait-and-scan versus active treatment). Such black-box-like solutions are widely accepted in practice nowadays, especially in connection with deep learning applications⁵⁵. Moreover, the user can still interact, even with the nontransparent classifiers, and analyze their

responses by manually adjusted inputs. An optimal classifier was selected as the best performing combination of a feature selection technique with a learned classifier. For the CBR data, it was found to be the combination of gradient boosting and a neural network; in the case of the PDA data set, the optimal performance was achieved using combinations of a logistic regression/neural network, or decision tree/logistic regression, or gradient boosting/logistic regression.

Potential limitations of our study and future directions. We are aware of the potential limitations of our study. Firstly, although we have assembled a relatively large amount of data from our participants, the final cleaned set contained a smaller number of records due to inconsistency in examination over the years (especially in the cases of ABR and OAE, as they were often not present during the initial examination), and unavailability of some variables in some of the records. A lower number of records may cause a decreased performance of the model, yet it avoids biases resulting from the usage of incomplete or potentially incorrect data. In the current analyses, we primarily focused on audiometric data, although information about potential vestibular pathology could be added to the decision making process in the future. Secondly, we omitted the patients' subjective input to avoid any subjectivity in the data set; however, our clinical experience shows that the subjective worsening of symptoms (that does not necessarily match the objective measurements) might be a strong factor influencing the decision about further VS treatment. Thirdly, our approach to the VS treatment is not purely based on objective measures, but also on the patients' preference and expectations, and also on the surgeons' experience and skill level; therefore the presented model is not expected to replace those inputs, but to support the decision making in deciding whether to directly opt for surgery or wait and scan.

Based on our results the future perspectives of our research using the supervised machine learning approach will be the inclusion of not only audiometric but also the vestibular data from our subjects, which would lead to an even more complex prediction model of the VS behavior. The conclusions formulated from supervised learning will be further enhanced with unsupervised analyses, including the linear and nonlinear clustering of data and variables, applied to the full-dimensional data set.

Conclusions

Using semi-supervised machine-learning algorithms complemented with expert (manual) interactive analyses, we developed practical tools to support the decision process related to the treatment of vestibular schwannomas. These tools comprise of simple decision rules (decision trees) for both static and dynamic data offering accuracy of around 80%, and automated black-box classifiers offering even better performance. Our results already indicate that from the initial data obtained at diagnosis (size of the tumor (Koo's classification and 1D size in T2 weighted MRI), speech perception (described by SRT) and pure tone average), it is possible to predict the need of VS active treatment. Furthermore, we propose minimum sets of diagnostic variables which are crucial for deciding on VS treatment. Overall, these findings can be used to make the diagnostic and decision-making procedures more time- and cost-efficient, by focusing on the important metrics and eliminating the unnecessary measurements.

Data availability

Data are available at the authors upon request.

Received: 1 June 2021; Accepted: 25 August 2021

Published online: 15 September 2021

References

- Springborg, J. B., Poulsen, L. & Thomsen, J. Nonvestibular schwannoma tumors in the cerebellopontine angle: A structured approach and management guidelines. *Skull Base* **18**, 217–227 (2008).
- Halliday, J., Rutherford, S. A., McCabe, M. G. & Evans, D. G. An update on the diagnosis and treatment of vestibular schwannoma. *Expert Rev. Neurother.* **18**, 29–39 (2018).
- Lee, J. D., Lee, B. D. & Hwang, S. C. Vestibular schwannoma in patients with sudden sensorineural hearing loss. *Skull Base* **21**, 75–78 (2011).
- Chovanec, M. *et al.* Does attempt at hearing preservation microsurgery of vestibular schwannoma affect postoperative tinnitus?. *BioMed Res. Int.* **2015**, 783169 (2015).
- Čada, Z. *et al.* Vertigo perception and quality of life in patients after surgical treatment of vestibular schwannoma with pretreatment prehabilitation by chemical vestibular ablation. *BioMed Res. Int.* **2016**, 11 (2016).
- Betka, J. *et al.* Complications of microsurgery of vestibular schwannoma. *BioMed Res. Int.* **2014**, 315952 (2014).
- Darrouzet, V., Martel, J., Enée, V., Bébér, J.-P. & Guérin, J. Vestibular schwannoma surgery outcomes: Our multidisciplinary experience in 400 cases over 17 years. *Laryngoscope* **114**, 681–688 (2004).
- Starnoni, D. *et al.* Surgical management for large vestibular schwannomas: A systematic review, meta-analysis, and consensus statement on behalf of the EANS skull base section. *Acta Neurochir. (Wien)* **162**, 2595–2617 (2020).
- Prasad, S. C. *et al.* Decision making in the wait-and-scan approach for vestibular schwannomas: Is there a price to pay in terms of hearing, facial nerve, and overall outcomes?. *Neurosurgery* **83**, 858–870 (2018).
- Cha, D., Shin, S. H., Kim, S. H., Choi, J. Y. & Moon, I. S. Machine learning approach for prediction of hearing preservation in vestibular schwannoma surgery. *Sci. Rep.* **10**, 7136 (2020).
- Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. From data mining to knowledge discovery in databases. *AI Mag.* **17**, 37–37 (1996).
- Chen, M.-S., Han, J. & Yu, P. S. Data mining: An overview from a database perspective. *IEEE Trans. Knowl. Data Eng.* **8**, 866–883 (1996).
- Patel, H. & Prajapati, P. Study and analysis of decision tree based classification algorithms. *Int. J. Comput. Sci. Eng.* **6**, 74–78 (2018).
- Quinlan, J. R. Induction of decision trees. *Mach. Learn.* **1**, 81–106 (1986).
- Sharma, H. & Kumar, S. A survey on decision tree algorithms of classification in data mining. *Int. J. Sci. Res. IJSR* **5**, 2094 (2016).
- Jin, X., Xu, A., Bie, R. & Guo, P. Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles. in *Data Mining for Biomedical Applications* (eds. Li, J., Yang, Q. & Tan, A.-H.) 106–115. https://doi.org/10.1007/11691730_11 (Springer, 2006).

17. Ren, Q., Cheng, H. & Han, H. Research on machine learning framework based on random forest algorithm. *AIP Conf. Proc.* **1820**, 080020 (2017).
18. Tin Kam Ho. Random decision forests. in *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Vol. 1. 278–282 (1995).
19. Laber, E. & Murtinho, L. Minimization of Gini impurity: NP-completeness and approximation algorithm via connections with the k-means problem. *Electron. Notes Theor. Comput. Sci.* **346**, 567–576 (2019).
20. Raileanu, L. & Stoffel, K. Theoretical comparison between the Gini index and information gain criteria. *Ann. Math. Artif. Intell.* **41**, 77–93 (2004).
21. Natekin, A. & Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurobot.* **7**, 11 (2013).
22. Zhang, Z., Zhao, Y., Canes, A., Steinberg, D. & Lyashevskaya, O. Predictive analytics with gradient boosting in clinical medicine. *Ann. Transl. Med.* **7**, 7 (2019).
23. Levy, J. J. & O'Malley, A. J. Don't dismiss logistic regression: The case for sensible extraction of interactions in the era of machine learning. *BMC Med. Res. Methodol.* **20**, 1–15 (2020).
24. Menard, S. W. *Logistic Regression: From Introductory to Advanced Concepts and Applications*. (SAGE, 2010).
25. Peng, J., Lee, K. & Ingersoll, G. An introduction to logistic regression analysis and reporting. *J. Educ. Res.* **96**, 3–14 (2002).
26. Draper, N. R. & Smith, H. *Applied Regression Analysis*. (Wiley-Interscience, 1998).
27. Chordia, T., Goyal, A. & Tong, Q. Pairwise correlations. *SSRN Electron. J.* <https://doi.org/10.2139/ssrn.1785390> (2011).
28. Meng, J. *et al.* Prognostic value of an immunohistochemical signature in patients with esophageal squamous cell carcinoma undergoing radical esophagectomy. *Mol. Oncol.* **12**, 196 (2017).
29. Muthukrishnan, R. & Rohini, R. LASSO: A feature selection technique in predictive modeling for machine learning. in *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*. 18–20. <https://doi.org/10.1109/ICACA.2016.7887916> (2016).
30. Battineni, G., Chintalapudi, N. & Amenta, F. Machine learning in medicine: Performance calculation of dementia prediction by support vector machines (SVM). *Inform. Med. Unlocked* **16**, 100200 (2019).
31. Cervantes, J., Garcia-Lamont, F., Rodriguez-Mazahua, L. & Lopez, A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* **408**, 189–215 (2020).
32. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
33. Amato, F. *et al.* Artificial neural networks in medical diagnosis. *J. Appl. Biomed.* **11**, 47–58 (2013).
34. Shahid, N., Rappon, T. & Berta, W. Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PLoS ONE* **14**, e212356 (2019).
35. Hajian-Tilaki, K. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Casp. J. Intern. Med.* **4**, 627–635 (2013).
36. ZouKelly, H., James, O. A. & Laura, M. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* **115**, 654–657 (2007).
37. Ebuwa, S. H., Sharif, M. S., Alazab, M. & Al-Nemrat, A. Variance ranking attributes selection techniques for binary classification problem in imbalance data. *IEEE Access* **7**, 24649–24666 (2019).
38. Beenstock, M. Predicting the stability and growth of acoustic neuromas. *Otol. Neurotol.* **23**, 542–549 (2002).
39. Artz, J. C. J. M., Timmer, F. C. A., Mulder, J. J. S., Cremers, C. W. R. J. & Graamans, K. Predictors of future growth of sporadic vestibular schwannomas obtained by history and radiologic assessment of the tumor. *Eur. Arch. Otorhinolaryngol.* **266**, 641–646 (2009).
40. Malhotra, P. S. *et al.* Clinical, radiographic, and audiometric predictors in conservative management of vestibular schwannoma. *Otol. Neurotol.* **30**, 507–514 (2009).
41. Agrawal, Y., Clark, J. H., Limb, C. J., Niparko, J. K. & Francis, H. W. Predictors of vestibular schwannoma growth and clinical implications. *Otol. Neurotol.* **31**, 807–812 (2010).
42. Timmer, F. C. A. *et al.* Prediction of vestibular schwannoma growth: A novel rule based on clinical symptomatology. *Ann. Otol. Rhinol. Laryngol.* **120**, 807–813 (2011).
43. Jethanamest, D. *et al.* Conservative management of vestibular schwannoma: Predictors of growth and hearing. *Laryngoscope* **125**, 2163–2168 (2015).
44. Hunter, J. B. *et al.* Single institutional experience with observing 564 vestibular schwannomas: Factors associated with tumor growth. *Otol. Neurotol.* **37**, 1630–1636 (2016).
45. D'Haese, S. *et al.* Vestibular schwannoma: Natural growth and possible predictive factors. *Acta Otolaryngol. (Stockh.)* **139**, 753–758 (2019).
46. Fieux, M. *et al.* MRI monitoring of small and medium-sized vestibular schwannomas: Predictors of growth. *Acta Otolaryngol. (Stockh.)* **140**, 361–365 (2020).
47. Kleijwegt, M., Bettink, F., Malessy, M., Putter, H. & Vandermeij, A. Clinical predictors leading to change of initial conservative treatment of 836 vestibular schwannomas. *J. Neurol. Surg. Part B Skull Base* **81**, 15–21 (2020).
48. Hentschel, M. A. *et al.* Development of a model to predict vestibular schwannoma growth: An opportunity to introduce new wait and scan strategies. *Clin. Otolaryngol.* **46**, 273–283 (2021).
49. Herwadker, A., Vokurka, E. A., Evans, D. G. R., Ramsden, R. T. & Jackson, A. Size and growth rate of sporadic vestibular schwannoma: Predictive value of information available at presentation. *Otol. Neurotol.* **26**, 86–92 (2005).
50. Wolbers, J. G. *et al.* Identifying at diagnosis the vestibular schwannomas at low risk of growth in a long-term retrospective cohort. *Clin. Otolaryngol.* **41**, 788–792 (2016).
51. Bellman, R. E. *Dynamic Programming*. (Princeton University Press, 1957).
52. Keogh, E. & Mueen, A. Curse of dimensionality. in *Encyclopedia of Machine Learning and Data Mining* (eds. Sammut, C. & Webb, G. I.) 314–315. https://doi.org/10.1007/978-1-4899-7687-1_192 (Springer, 2017).
53. Venkat, N. *The Curse of Dimensionality: Inside Out*. <https://doi.org/10.13140/RG.2.2.29631.36006> (2018).
54. Verleysen, M. & François, D. The curse of dimensionality in data mining and time series prediction. in *Computational Intelligence and Bioinspired Systems* (eds. Cabestany, J., Prieto, A. & Sandoval, F.) Vol. 3512. 758–770 (Springer, 2005).
55. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

Acknowledgements

This work was supported by Czech Science Foundation (Grantová agentura České republiky, GAČR), project “Changes in the auditory cortex in patients with single sided deafness”, reg. no. 19-08241S.

Author contributions

Study design and supervision—O.P., M.C. & J.B.; data acquisition and pre-processing—Zu.B., J.B. & Z.F.; data analysis—J.V. & Zb.B.; manuscript preparation—Zb.B., J.V. & O.P. All authors discussed the results and commented on the manuscript during manuscript preparation.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Z.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021