

Predicting surgical decision-making in vestibular schwannoma using tree-based machine learning

Ron Gadot, BSc,¹ Adrish Anand, BA,¹ Benjamin D. Lovin, MD,² Alex D. Sweeney, MD,² and Akash J. Patel, MD^{1,3}

¹Department of Neurosurgery, Baylor College of Medicine; ²Department of Otolaryngology-Head and Neck Surgery, Baylor College of Medicine, Houston; and ³Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, Texas

OBJECTIVE Vestibular schwannomas (VSs) are the most common neoplasm of the cerebellopontine angle in adults. Though these lesions are generally slow growing, their growth patterns and associated symptoms can be unpredictable, which may complicate the decision to pursue conservative management versus active intervention. Additionally, surgical decision-making can be controversial because of limited high-quality evidence and multiple quality-of-life considerations. Machine learning (ML) is a powerful tool that utilizes data sets to essentialize multidimensional clinical processes. In this study, the authors trained multiple tree-based ML algorithms to predict the decision for active treatment versus MRI surveillance of VS in a single institutional cohort. In doing so, they sought to assess which preoperative variables carried the most weight in driving the decision for intervention and could be used to guide future surgical decision-making through an evidence-based approach.

METHODS The authors reviewed the records of patients who had undergone evaluation by neurosurgery and otolaryngology with subsequent active treatment (resection or radiation) for unilateral VS in the period from 2009 to 2021, as well as those of patients who had been evaluated for VS and were managed conservatively throughout 2021. Clinical presentation, radiographic data, and management plans were abstracted from each patient record from the time of first evaluation until the last follow-up or surgery. Each encounter with the patient was treated as an instance involving a management decision that depended on demographics, symptoms, and tumor profile. Decision tree and random forest classifiers were trained and tested to predict the decision for treatment versus imaging surveillance on the basis of unseen data using an 80/20 pseudorandom split. Predictor variables were tuned to maximize performance based on lowest Gini impurity indices. Model performance was optimized using fivefold cross-validation.

RESULTS One hundred twenty-four patients with 198 rendered decisions concerning management were included in the study. In the decision tree analysis, only a maximum tumor dimension threshold of 1.6 cm and progressive symptoms were required to predict the decision for treatment with 85% accuracy. Optimizing maximum dimension thresholds and including age at presentation boosted accuracy to 88%. Random forest analysis ($n = 500$ trees) predicted the decision for treatment with 80% accuracy. Factors with the highest variable importance based on multiple measures of importance, including mean minimal conditional depth and largest Gini impurity reduction, were maximum tumor dimension, age at presentation, Koos grade, and progressive symptoms at presentation.

CONCLUSIONS Tree-based ML was used to predict which factors drive the decision for active treatment of VS with 80%–88% accuracy. The most important factors were maximum tumor dimension, age at presentation, Koos grade, and progressive symptoms. These results can assist in surgical decision-making and patient counseling. They also demonstrate the power of ML algorithms in extracting useful insights from limited data sets.

<https://thejns.org/doi/abs/10.3171/2022.1.FOCUS21708>

KEYWORDS schwannoma; machine learning; decision tree; random forest; surgery

ABBREVIATIONS BSLMC = Baylor St. Luke's Medical Center; BTGH = Ben Taub General Hospital; CPA = cerebellopontine angle; IAC = intracanalicular; ML = machine learning; OOB = out of bag; RF = random forest; SRS = stereotactic radiosurgery; VS = vestibular schwannoma.

SUBMITTED November 22, 2021. **ACCEPTED** January 19, 2022.

INCLUDE WHEN CITING DOI: 10.3171/2022.1.FOCUS21708.

VESTIBULAR schwannomas (VSs) account for 8%–10% of all intracranial neoplasms and are the most common tumors of the cerebellopontine angle (CPA) in adults.¹ They are sporadic and unilateral in 95% of cases but can occur more frequently in genetic susceptibility syndromes such as neurofibromatosis type 2 and schwannomatosis.^{1,2} Though VSs are often considered rare, recent epidemiological studies have found an increasing lifetime prevalence of as many as 1 case in 500 people, likely because of increased access to neurodiagnostic imaging and earlier or incidental diagnoses.^{1,3,4} Current management strategies for VSs vary considerably across centers and countries, partially due to a complex interplay of uncertain clinical behavior and functional quality-of-life concerns.^{5,6} Symptoms of VS can be caused by compression or destruction of surrounding structures, inner ear toxicity, as well as cerebrospinal fluid flow obstruction, and the most common symptoms are hearing loss, tinnitus, and/or balance disorders.^{7,8} In principle, the decision that clinicians must make for these patients at each visit is whether to surveil a patient with serial MRI (wait and scan) or to intervene (whether through microsurgery or radiosurgery) to prevent or curtail neurological deterioration.

Tumor size parameters, including initial tumor size and interval growth on serial imaging (often > 2 mm between images), are commonly associated with the decision for surgery. Still, there are patient and provider-related factors that also drive decision-making.^{5,9} Experts from high-volume specialty centers have suggested experience-based thresholds for considering surgical intervention, that is, greater than a 1.5-cm maximum tumor diameter in the CPA.¹ Recently, evidence-based methods have been gaining popularity in guiding VS surgical decision-making with a particular focus on machine learning (ML) approaches.^{10,11} Still, treatment algorithms from diverse centers and populations are desired to discern which patient and tumor-related variables should reliably prompt active VS treatment. Tree-based ML provides a powerful set of tools to accomplish classification (categorical outcomes) or regression (continuous outcomes) tasks in multidimensional clinical settings.¹²

We trained multiple tree-based ML algorithms to predict the decision for active treatment versus MRI surveillance of VS in a single institutional cohort. In doing so, we sought to assess which preoperative variables carried the most weight in driving the decision for intervention and could be used to guide future surgical decision-making through an evidence-based approach.

Methods

Data Acquisition and Characteristics

We reviewed the records of patients who had undergone evaluation by neurosurgery and otolaryngology with subsequent treatment via microsurgical resection or stereotactic radiosurgery (SRS) for unilateral VS from 2009 to 2021 at two Baylor College of Medicine–affiliated hospitals. Additionally, we reviewed the records of patients with VS who had been seen in clinic throughout 2021 and had ultimately been managed conservatively. Patients were ex-

cluded if they had ICD-10 codes for a VS diagnosis but had no intracanalicular (IAC) or CPA component to their tumor. Each encounter with the patient was treated as an instance involving a management decision that depended on demographics, symptoms, and tumor profile through a case-based reasoning approach.¹¹ This study was conducted following the approval by the Baylor College of Medicine Institutional Review Board.

Fifteen clinical variables available at each visit including the end-of-encounter decision were abstracted onto a standardized chart. Demographic variables included hospital system of evaluation (Baylor St. Luke's Medical Center [BSLMC] or Ben Taub General Hospital [BTGH]), age at presentation, gender, tumor image or encounter number for that patient (first, second, etc.), and whether the patient had a prior history of treatment for the tumor. Tumor variables included tumor location (IAC, CPA, or IAC+CPA), maximum tumor dimension, presence of a fundal cap, Koos grade,¹³ whether the tumor did (Koos grades III–IV) or did not (Koos grades I–II) have brainstem involvement ("Koos_half"), whether the tumor resulted in brainstem compression (Koos grade IV) or did not (Koos grades I–III; "Koos_worst"), and presence of hydrocephalus. Tumor-related symptoms, including hearing loss, tinnitus, imbalance, vertigo, headaches, and facial pain/weakness, were reviewed at each encounter relative to any previous encounters and were grouped into any new and/or progressive symptoms at presentation. While several of these variables have high collinearity (i.e., Koos grade and brainstem involvement), we included these categorizations to emulate various decision-making workflows utilized by different providers in day-to-day practice.

Statistical Analysis

All statistical analyses were performed in RStudio statistical computing software (R Foundation for Statistical Computing) using rpart (RRID:SCR_021777), randomForest (RRID:SCR_015718), randomForestExplainer (RRID:021823), and gtsummary (RRID:021319) packages. Summary statistics were performed on all clinical variables, comparing encounters that resulted in MRI surveillance against those that resulted in active treatment. Chi-square or Fisher's exact tests and Wilcoxon rank-sum tests were used to ascertain associative differences between the two encounter groups, with a false discovery rate-adjusted p value < 0.05 considered to be statistically significant.

Tree-Based Modeling

Decision tree algorithms were trained and validated on unseen data using an 80/20 pseudorandom split. Decision trees grow from a root node by recursively partitioning a data set based on random predictor variables, repeatedly assessing error in classification using, among other methods, the Gini impurity index.^{14,15} Variables with the largest imparted Gini impurity reduction are prioritized in the algorithm, and subsequent nodes are branched from each node until terminal "leaf" nodes do not further increase classification accuracy or are manually constrained by the user. Parameter hyper-tuning was performed to maximize

accuracy. For random forest (RF) analysis, a “bagging” approach was used in which the algorithm recursively bootstraps the data set with $n = 500$ tree iterations to generate an aggregate classification model. At each iteration, the model uses data not included in the bootstrapped data set (out-of-bag [OOB] sample) and classifies the sample based on the incrementally growing mode of the training set’s output.¹⁶ An RF model’s performance can be assessed using the OOB error as well as its accuracy in correctly classifying unseen data.¹⁷ The minimum number of parameters assessed for splitting at each step is, by convention, the square root of the total number of predictor variables. Parameter tuning was achieved using fivefold cross validation.

Results

One hundred twenty-four patients with 198 total patient encounters were included in this study. One hundred seven encounters (54%) resulted in the decision for continued MRI surveillance and 91 (46%) resulted in the decision for active intervention. The mean number of patient encounters until the last follow-up or the decision for active treatment was 1.63 (range 1–5). Table 1 demonstrates all gathered clinical variables stratified by encounter decision. Gender, recurrent tumor status, and the presence of a fundal cap were not significantly associated with decision category. Patients offered treatment were significantly younger (mean age 48 ± 14 years) than those recommended for continued MRI surveillance (mean age 57 ± 15 years; $p < 0.001$). Maximum tumor dimension was significantly larger for patients offered treatment (2.47 ± 1.05 cm) than for those offered serial imaging (1.07 ± 0.59 cm; $p < 0.001$). Patients in encounters that had resulted in treatment were significantly more likely to report new tumor-related symptoms (60%) than those in encounters that had resulted in MRI surveillance (21%; $p < 0.001$). Patient encounters at BTGH were more likely to result in surgery than those at BSLMC, likely given that BTGH is our institution’s county hospital where patients tended to present with substantially larger tumors (3.03 ± 1.26 vs 1.56 ± 0.98 cm, $p < 0.001$) causing advanced compressive or obstructive pathologies. Other variables noted to be significantly different in decision status included image number, tumor location, Koos grade, brainstem involvement, brainstem compression, and progressive symptoms at presentation.

Decision Tree Analysis

A decision tree was trained on approximately 80% of the available data (Fig. 1A). A maximum tumor dimension > 1.6 cm was found in 47% of encounters and conferred a 78% chance of being classified for treatment. Among those encounters, patients who had a maximum tumor dimension > 2.3 cm had a 95% chance of being offered treatment, whereas those with a maximum dimension of 1.6–2.3 cm were next stratified by the presence of progressive symptoms at presentation whereby those with progressive symptoms and a maximum tumor size between 1.6 and 2.3 cm had an 81% chance of being classified for treatment. The accuracy of this model in predicting the need for treatment on unseen test data was 85% (Fig. 1B).

TABLE 1. Baseline demographic, tumor, and symptom variables at 198 encounters

Variable	MRI Surveillance	Active Treatment	Adjusted p Value*
No. of encounters	107 (54)	91 (46)	0.005
Hospital			
BSLMC	102/107 (95)	75/91 (82)	
BTGH	5/107 (5)	16/91 (18)	
Age in yrs	57 ± 15	48 ± 14	<0.001
Gender			0.90
M	43/107 (40)	38/91 (42)	
F	64/107 (60)	53/91 (58)	
Image/encounter no.			0.015
1st	56/107 (52)	67/91 (74)	
2nd	27/107 (25)	13/91 (14)	
3rd	17/107 (16)	5/91 (5)	
4th	6/107 (6)	4/91 (4)	
5th	1/107 (1)	2/91 (2)	
Max tumor dimension in cm	1.07 ± 0.59	2.47 ± 1.05	<0.001
Recurrent tumor	7/107 (7)	11/91 (12)	0.20
Location			<0.001
IAC+CPA	37/107 (35)	68/91 (75)	
IAC	56/107 (52)	9/91 (10)	
CPA	14/107 (13)	14/91 (15)	
Fundal cap	26/107 (24)	68/91 (75)	0.76
Hydrocephalus	1/107 (1)	14/91 (15)	<0.001
Koos grade			<0.001
I	55/107 (51)	7/91 (8)	
II	34/107 (32)	15/91 (16)	
III	16/107 (15)	26/91 (29)	
IV	2/107 (2)	43/91 (47)	
Brainstem involvement			<0.001
Koos grades I–II	89/107 (83)	23/91 (25)	
Koos grades III–IV	18/107 (17)	68/91 (75)	
Brainstem compression			<0.001
Koos grades I–III	105/107 (98)	43/91 (47)	
Koos grade IV	2/107 (2)	48/91 (53)	
Progressive symptoms	12/107 (11)	29/91 (32)	<0.001
New symptoms	22/107 (21)	55/91 (60)	<0.001

Values represent number (%) or mean \pm standard deviation, unless indicated otherwise.

* Wilcoxon rank-sum test was used for continuous variables and Pearson’s chi-square or Fisher’s exact test was used for categorical variables as appropriate. The false discovery rate-adjusted p value is reported. Boldface type indicates statistical significance.

Classification fidelity was manually verified for optimal accuracy by tuning of tree minimum splitting, minimum bucket observation, and maximum depth values. Further model splitting of tumors < 1.6 cm by progressive symptoms and a < 1.3 -cm threshold as well as splitting tumors > 1.6 cm further by a maximum dimension > 2.3 cm and age at presentation boosted classification accuracy to 88% on unseen test data (Fig. 1C and D).

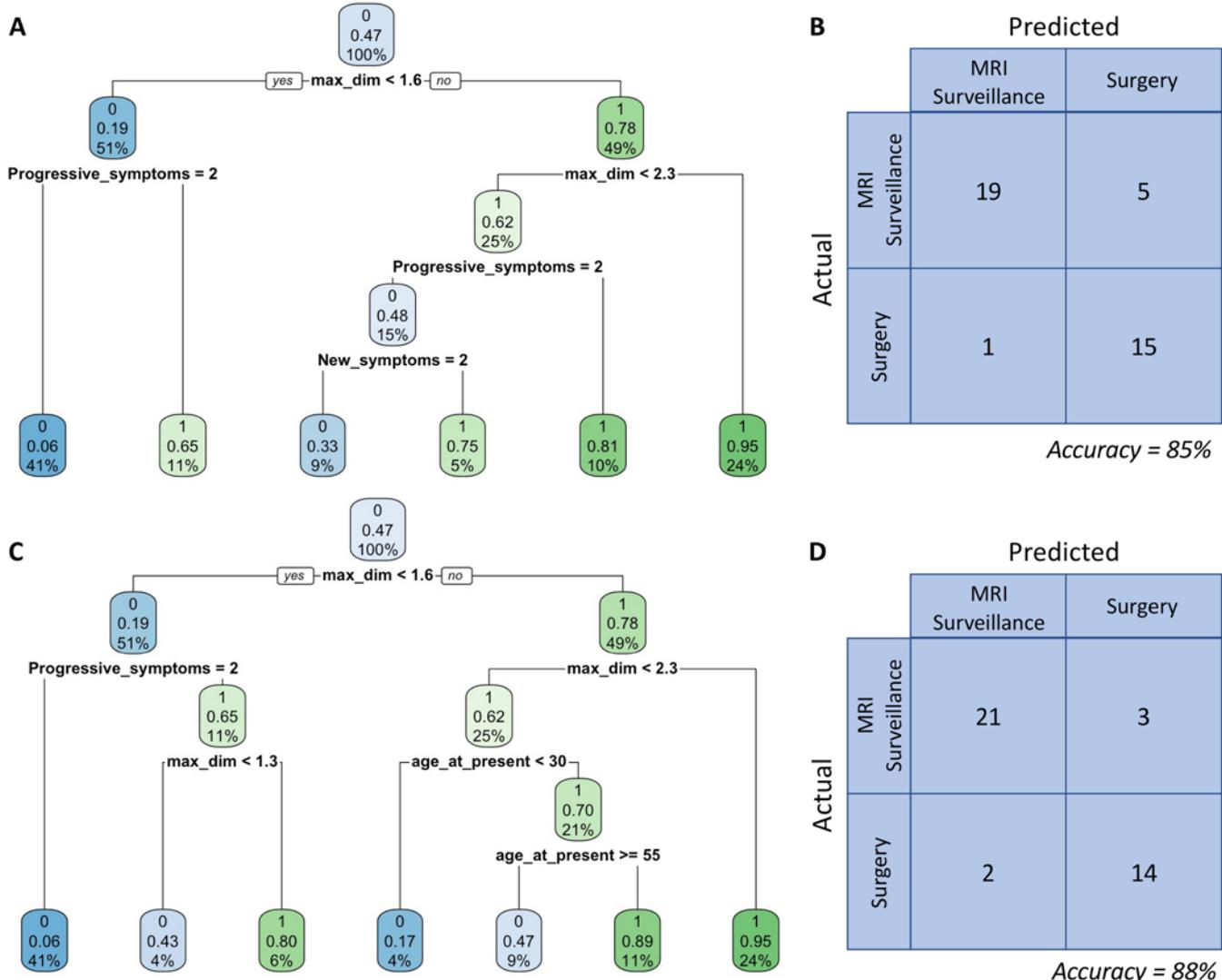


FIG. 1. A: Decision tree generated by an algorithm using default settings. A value of 0 was coded for MRI surveillance, whereas 1 denoted treatment. For example, the leftmost terminal leaf shows that the model assigned a 6% chance for active treatment if a patient had a maximum (max) tumor dimension (dim) < 1.6 cm and no symptom progression, which occurred in 41% of encounters. If a patient had a maximum tumor dimension of 1.6–2.3 cm but no progressive or new symptoms (new/progressive symptoms = 2) they had a 33% chance for surgery. B: Confusion matrix for unseen test data indicated 85% accuracy in the classification based on the model in panel A. C: Tuned decision tree that optimized accuracy in classification. D: Confusion matrix on unseen data indicated an 88% accuracy in the classification based on the adjusted model in panel C.

RF Analysis

An RF of $n = 500$ trees with a minimum of 3 variables assessed per node was trained on approximately 80% of the available data. The OOB sample error rate was 17.72%. The learning curve for the RF is depicted in Fig. 2. The model demonstrated a 16% error rate in misclassifying MRI surveillance as treatment and an 18% error rate in misclassifying treatment as MRI surveillance in the OOB sample. This indicates that the model was slightly better at detecting when a case demonstrated a need for treatment than it was at determining when MRI surveillance could be continued (higher sensitivity, lower specificity for treatment). Classification accuracy stabilized by approximately 200 trees grown. When tested on unseen data, the RF

model performed with 80% accuracy with an area under the curve of 0.83.

To identify which factors were most useful in accurate classification, the distribution of the minimum variable depth was plotted across all trees (Fig. 3). The mean minimal depth corresponds to the average minimal distance between the variable and the root node. Lower values of the minimal depth indicate greater importance in decision-making. Maximum tumor dimension (mean minimal depth = 1.83), age at presentation (2.3), and Koos grade (2.33) were the 3 most important variables used for splitting. Multiway variable importance plots demonstrate the relative importance of each of the top 10 most important variables based on three measures of variable importance

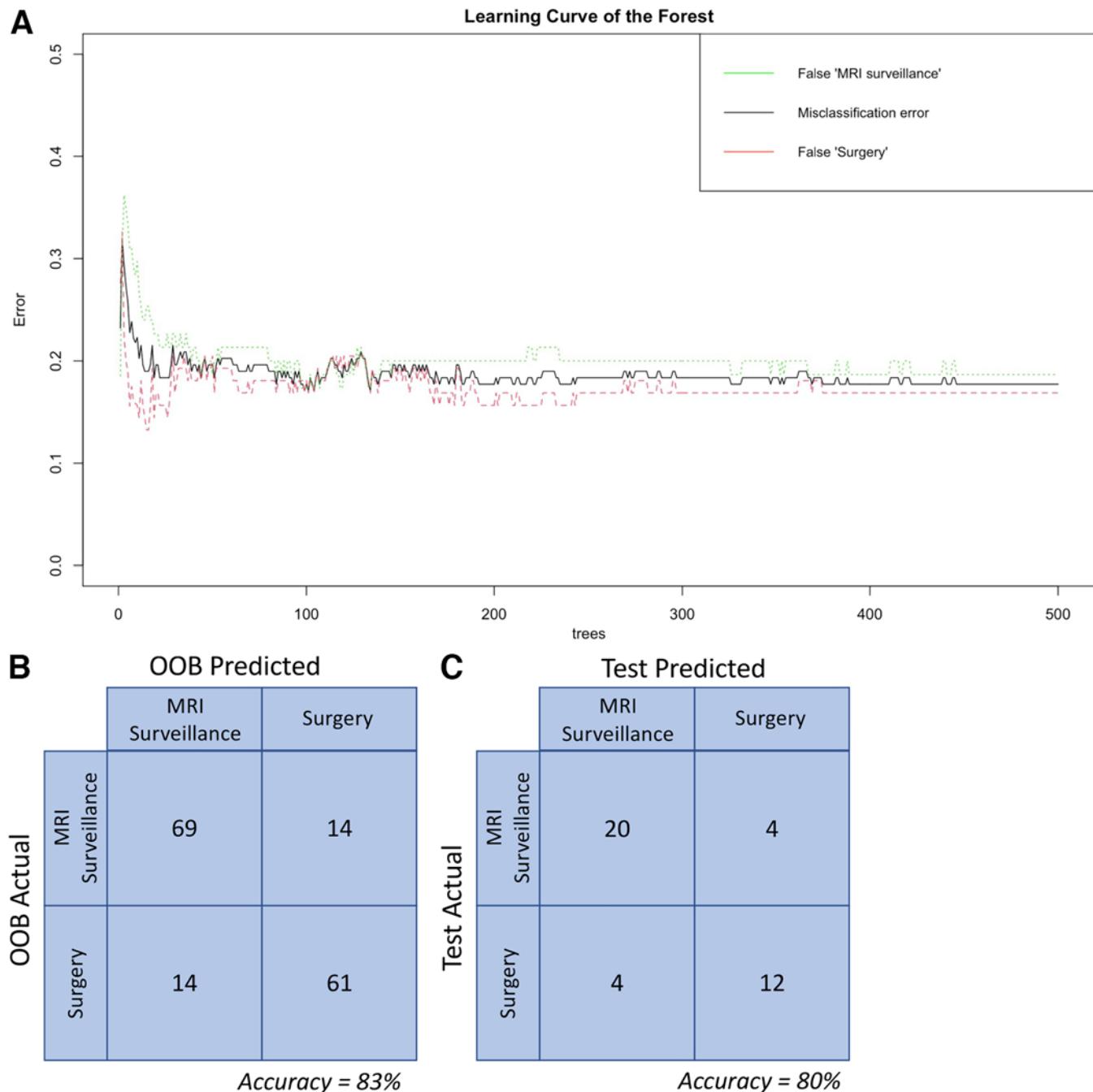


FIG. 2. RF analysis. Model learning curve (A) demonstrated with the error in OOB sample classification with increasing tree growth. Error rates stabilize after approximately 200 trees grown. Overall misclassification error on the training set was found to be 17.72% and was slightly higher at misclassifying cases that had received treatment than those that had received MRI surveillance. Confusion matrices showing OOB sample accuracy (83%, B) and unseen test data prediction accuracy (80%, C).

(Fig. 4A) as well as the role that each variable played in prediction (Fig. 4B).

Variable interactions, a metric for the proximity between two variables in a tree, were assessed using conditional minimal depth. This is calculated on the basis of the minimal depth of a second variable y in a tree in which a first variable x is a root and is calculated across all trees (Fig. 5). This analysis showed that the interactions

among maximum tumor dimension, age at presentation, and Koos grade occurred most frequently and within closest proximity to each other in the encounter stratification. Model classification predictions based on the interaction among these 3 most conditionally important variables were plotted (Fig. 6). It was evident that the model consistently classified an encounter for treatment in tumors with a maximum dimension of approximately $> 1.6\text{--}1.8$

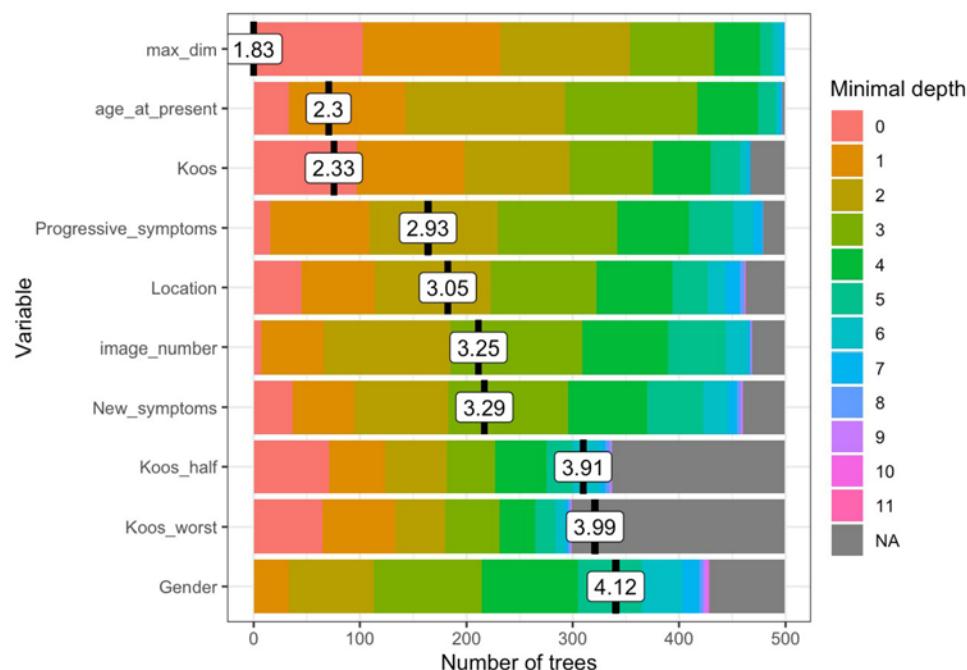


FIG. 3. The minimal depth of a variable across all tree nodes is plotted, with means designated by vertical bars color coded by frequency of depth. The lower the mean minimal depth, the more often that variable was used higher in a tree for splitting and the more purity that variable contributes to accurate classification. dim = dimension; Koos_half = Koos grade I or II versus III or IV; Koos_worst = Koos grade IV versus I–III; present = presentation.

cm, especially in patients between 35 and 55 years of age (Fig. 6A). This prediction trend held true for Koos grade III–IV tumors and a similar age range (Fig. 6B). Finally, the predictive interaction between Koos grade and maximum tumor dimension were plotted, showing that between the two variables, a maximum dimension of slightly less than 2 cm or larger often drove the decision for treatment

more than a higher Koos grade status at any maximum tumor dimension.

Discussion

In this study, we sought to demonstrate the utility of ML tools in predicting surgical decision-making in VS.

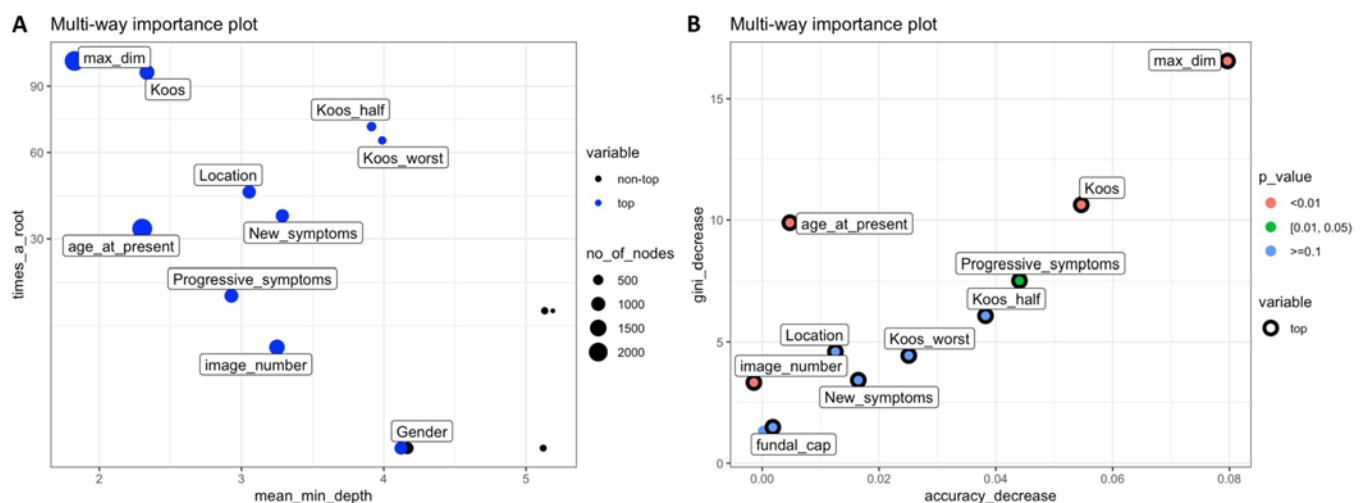


FIG. 4. **A:** Multiway importance plot demonstrating three measures of variable importance: mean depth of first split, number of trees in which the root is split on the variable, and total number of nodes in the forest that split on that variable. “Top” indicates whether the variable was in the top 10 most important variables. **B:** Multiway importance plot demonstrating variable predictive ability. Gini impurity decrease in the variable’s presence is regressed against classification accuracy decrease in the variable’s absence during splitting. The p values are derived from the model’s distribution of splitting based on the variable compared to a binomial distribution of the number of nodes split on the variable assuming that variables are randomly drawn to form splits.

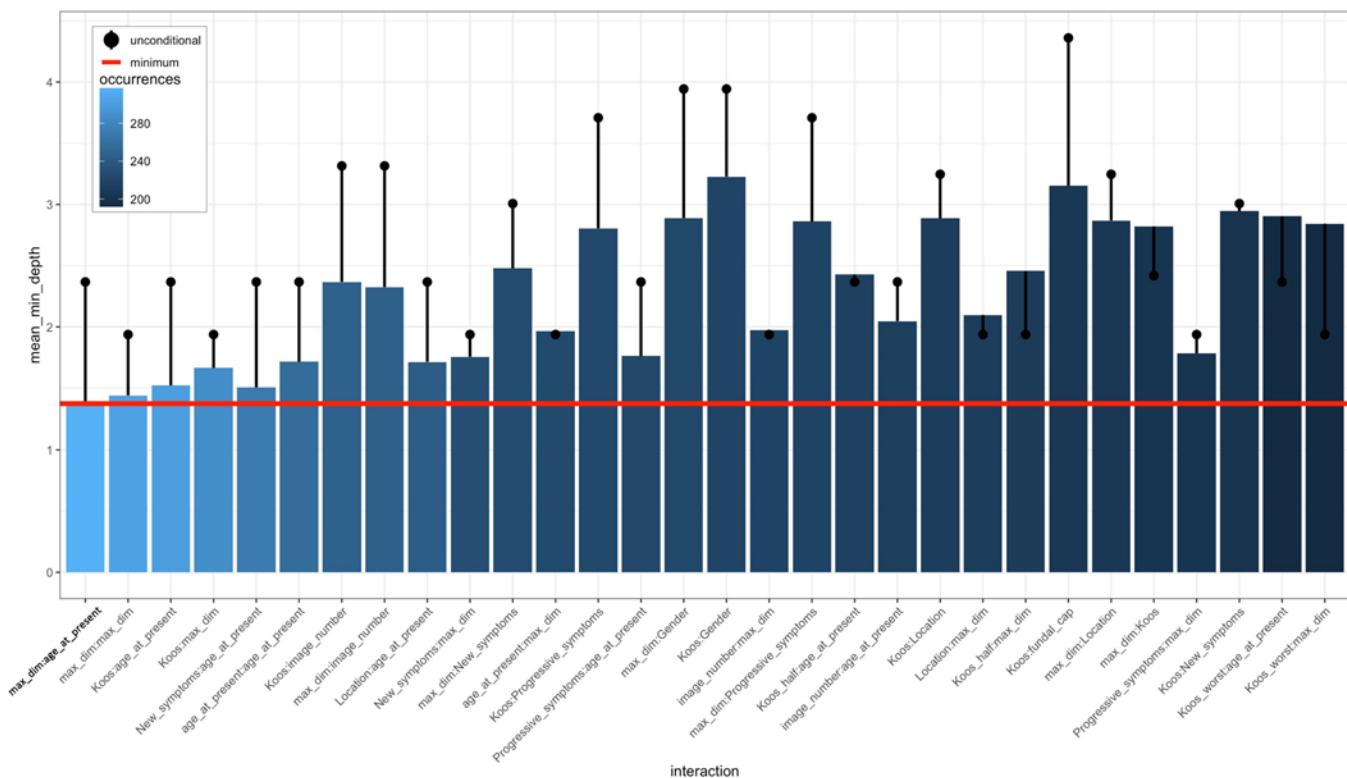


FIG. 5. The conditional minimal depth plot depicts the 30 interactions that appeared most frequently in the RF analysis. The red horizontal line depicts the minimum value for the minimal mean conditional depth, which occurred between maximum (max) tumor dimension (dim) and age at presentation (present).

Our results not only help to understand clinical practice at our institution, but may also help to guide and complement future practice at our and other institutions. Using tree-based ML approaches including decision tree and RF analyses, we trained algorithms to predict a surgical provider's decision for MRI surveillance or active treatment with 80%–88% accuracy.

The most important variables that could predict a decision for active management were maximum tumor dimension, age at presentation, Koos grade, and progressive symptom development. Specifically, a maximum tumor dimension of 1.6–1.8 cm was found to be a convergent threshold prompting treatment regardless of the other clinical variables assessed. This finding is in line with recently published results involving a large cohort from a high-volume center in which the optimal tumor size thresholds associated with various improved clinical outcomes in VS surgery were ascertained.¹⁰ Authors of that study used logistic regression models to determine that tumors within 3 mm of a 17-mm maximum tumor size threshold were more amenable to a resection that resulted in better outcomes for their patients.¹⁰ Although evaluating our surgical cohort's outcomes was outside the scope of this work, we are assured to see that our active treatment–prompting tumor size threshold lines up with other centers in guiding our institution's practice.

One other group recently used ML approaches to predict the need for active treatment in VS. Using a similar case-based reasoning method, the authors found that Koos

classification, speech reception threshold, and pure tone audiometry predicted the need for active treatment with around 80% accuracy.¹¹ More generally, ML is gaining popularity in the surgical decision-making literature at large, with multiple published reviews amassing evidence for its use in gaining novel insights into hospitalization needs, intensive care unit stays, surgical complications, and surgical decision aids.^{18,19}

ML is a branch of artificial intelligence in which computers are given data and allowed to freely learn from information. ML methods can be categorized into supervised and unsupervised learning, which can each be subdivided into several statistical and computational approaches. Tree-based ML is one category of supervised learning in which classification or regression questions can be answered on the basis of error reduction techniques.¹² Decision trees are simple yet powerful tools for essentializing interactions in a multidimensional data set but can be prone to overfitting.¹⁴ This problem is improved by setting constraints on model parameters and pruning nodes that do not increase variance but may contribute to model bias.²⁰ Methods to address the trade-off between bias and variance in growing decision trees include data bagging, boosting, and stacking, which are performed by various types of ensemble learning.²¹

RFs are a widely used form of ensemble learning, owing their popularity to strong multidimensionality reduction capabilities and within-training-set internal validation.^{22,23} One concern in the use of RFs in medical studies is that

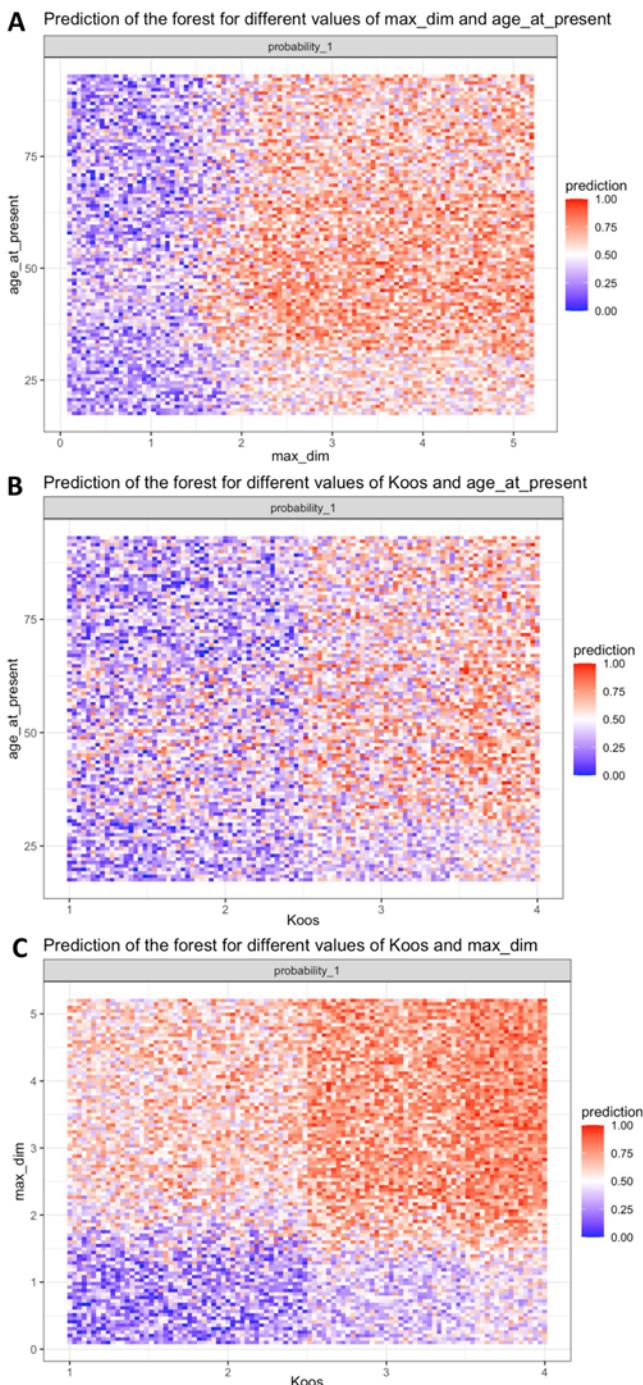


FIG. 6. Prediction scaled from MRI surveillance (0) to active treatment (1) based on interaction between maximum tumor dimension and age at presentation (A), Koos grade (indicated in Arabic numerals in this figure) and age at presentation (B), and maximum tumor dimension and Koos grade (C).

the forest can be a “black box,” as no unified decision tree algorithm can be visually digested when attempting to sort through clinical problems. This concern is mitigated by assessing variable importance, as was done in our study. Further, RFs, as opposed to other statistical approaches such as logistic regression, may be better suited for our clinical

question, as they have been shown to perform better in data sets with high nonlinearity and multifactorial relationships between dependent and independent variables.²⁴

Our study has several limitations. Most importantly, it is a retrospective study involving a single institutional cohort and is therefore limited by our institutional practices over the study period. Our findings would be further bolstered by test validation on an independent external cohort assessing similar basic clinical variables as in our model. We have provided our raw data set (Supplemental Table 1) for transparency and reproducibility, as sharing data in the literature is key to the growing utilization of ML in medical and scientific study.¹⁹ The relatively limited sample size is overcome by RF methods of internal cross-validation that decrease overfitting of the data.²² Additionally, our study involves the largest patient cohort used in tree-based VS surgical decision-making prediction to date.¹¹ Patients who underwent resection or SRS were grouped into an active treatment category, as there were significantly fewer SRS treatment decisions ($n = 11$) compared with decisions for microsurgery ($n = 80$) or MRI surveillance ($n = 107$) for an adequate three-way comparison. Future work should involve an expanded analysis with a larger sample of patients who undergo SRS to ascertain potential differences in factors driving microsurgery versus radiosurgery in the “gray area” of VS cases (i.e., tumors between 1.6 and 2.3 cm or older patients with relative contraindications to surgery). Finally, it is important to note that possibly many other variables including patient, surgeon, and co-medical factors may have further influenced decision-making at any given encounter. We attempted to use some of the most commonly thought about factors that drive surgical decision-making in day-to-day practice to build an evidence-based model with which to guide best practice. These results may ultimately assist in simplifying VS treatment strategies, reducing the number of specialist examinations, and increasing cost effectiveness.¹¹

Conclusions

Using tree-based ML tools, we trained classifiers that could predict which factors drive the decision for active treatment in patients with VS with an accuracy of 80%–88%. The most important factors were maximum tumor dimension, age at presentation, Koos grade, and progressive symptom development. In addition to highlighting practice patterns, these findings can assist in surgical decision-making and patient counseling. The results also showcase the power of ML algorithms in extracting useful data from even limited data sets.

Acknowledgments

We thank the patients without whom this study would not have been possible.

References

- Carlson ML, Link MJ. Vestibular schwannomas. *N Engl J Med.* 2021;384(14):1335-1348.
- Mohyuddin A, Neary WJ, Wallace A, et al. Molecular genetic analysis of the NF2 gene in young patients with unilateral vestibular schwannomas. *J Med Genet.* 2002;39(5):315-322.

3. Marinelli JP, Lohse CM, Carlson ML. Incidence of vestibular schwannoma over the past half-century: a population-based study of Olmsted County, Minnesota. *Otolaryngol Head Neck Surg*. 2018;159(4):717-723.
4. Reznitsky M, Petersen MMBS, West N, Stangerup SE, Cayé-Thomasen P. Epidemiology of vestibular schwannomas - prospective 40-year data from an unselected national cohort. *Clin Epidemiol*. 2019;11:981-986.
5. Carlson ML, Glasgow AE, Grossardt BR, Habermann EB, Link MJ. Does where you live influence how your vestibular schwannoma is managed? Examining geographical differences in vestibular schwannoma treatment across the United States. *J Neurooncol*. 2016;129(2):269-279.
6. Gauden A, Weir P, Hawthorne G, Kaye A. Systematic review of quality of life in the management of vestibular schwannoma. *J Clin Neurosci*. 2011;18(12):1573-1584.
7. Halliday J, Rutherford SA, McCabe MG, Evans DG. An update on the diagnosis and treatment of vestibular schwannoma. *Expert Rev Neurother*. 2018;18(1):29-39.
8. Dilwali S, Landegger LD, Soares VY, Deschler DG, Stankovic KM. Secreted factors from human vestibular schwannomas can cause cochlear damage. *Sci Rep*. 2015;5:18599.
9. Carlson ML, Tveiten ØV, Lund-Johansen M, Tombers NM, Lohse CM, Link MJ. Patient motivation and long-term satisfaction with treatment choice in vestibular schwannoma. *World Neurosurg*. 2018;114:e1245-e1252.
10. Macielak RJ, Wallerius KP, Lawlor SK, et al. Defining clinically significant tumor size in vestibular schwannoma to inform timing of microsurgery during wait-and-scan management: moving beyond minimum detectable growth. *J Neurosurg*. Published online October 15, 2021. doi: 10.3171/2021.4.JNS21465
11. Profant O, Bureš Z, Balogová Z, et al. Decision making on vestibular schwannoma treatment: predictions based on machine-learning analysis. *Sci Rep*. 2021;11(1):18376.
12. Banerjee M, Reynolds E, Andersson HB, Nallamothu BK. Tree-based analysis. *Circ Cardiovasc Qual Outcomes*. 2019; 12(5):e004879.
13. Erickson NJ, Schmalz PGR, Agee BS, et al. Koos classification of vestibular schwannomas: a reliability study. *Neurosurgery*. 2019;85(3):409-414.
14. Song YY, Lu Y. Decision tree methods: applications for classification and prediction. *Shanghai Jingshen Yixue*. 2015; 27(2):130-135.
15. Nembrini S, König IR, Wright MN. The revival of the Gini importance? *Bioinformatics*. 2018;34(21):3711-3718.
16. Svetnik V, Liaw A, Tong C, Culberson JC, Sheridan RP, Feuston BP. Random forest: a classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comput Sci*. 2003;43(6):1947-1958.
17. Breiman L. Random forests. *Mach Learn*. 2001;45:5-32.
18. Byerly S, Maurer LR, Mantero A, Naar L, An G, Kaafarani HMA. Machine learning and artificial intelligence for surgical decision making. *Surg Infect (Larchmt)*. 2021;22(6): 626-634.
19. Loftus TJ, Tighe PJ, Filiberto AC, et al. Artificial intelligence and surgical decision-making. *JAMA Surg*. 2020;155(2):148-158.
20. Ahmed AM, Rizaner A, Ulusoy AH. A novel decision tree classification based on post-pruning with Bayes minimum risk. *PLoS One*. 2018;13(4):e0194168.
21. Wang CW. New ensemble machine learning method for classification and prediction on gene expression data. *Conf Proc IEEE Eng Med Biol Soc*. 2006;2006:3478-3481.
22. Touw WG, Bayjanov JR, Overmars L, et al. Data mining in the life sciences with random forest: a walk in the park or lost in the jungle? *Brief Bioinform*. 2013;14(3):315-326.
23. Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. *Front Aging Neurosci*. 2017;9: 329.
24. Couronné R, Probst P, Boulesteix AL. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics*. 2018;19(1):270.

Disclosures

Dr. Sweeney reports being a consultant for Cochlear Americas, MED-EL GmbH, Oticon Medica GmbH, and Advanced Bionics Corp.

Author Contributions

Conception and design: Patel, Gadot. Acquisition of data: Gadot, Anand, Lovin. Analysis and interpretation of data: all authors. Drafting the article: Gadot. Critically revising the article: Patel, Gadot, Anand, Sweeney. Reviewed submitted version of manuscript: Patel, Gadot, Sweeney. Statistical analysis: Gadot. Administrative/technical/material support: Patel. Study supervision: Patel, Sweeney.

Supplemental Information

Online-Only Content

Supplemental material is available online.

Supplemental Table 1. <https://thejns.org/doi/suppl/10.3171/2022.1.FOCUS21708>.

Correspondence

Akash J. Patel: Baylor College of Medicine, Houston, TX. akash.patel@bcm.edu.