Neuro-Oncology

24(4), 601–609, 2022 | https://doi.org/10.1093/neuonc/noab211 | Advance Access date 6 September 2021

Machine learning approach to differentiation of peripheral schwannomas and neurofibromas: A multi-center study

Michael Zhang[†], Elizabeth Tong[†], Sam Wong, Forrest Hamrick, Maryam Mohammadzadeh, Vaishnavi Rao, Courtney Pendleton, Brandon W. Smith, Nicholas F. Hug, Sandip Biswal, Jayne Seekins, Sandy Napel, Robert J. Spinner, Mark A. Mahan, Kristen W. Yeom[†], and Thomas J. Wilson[†]

Department of Neurosurgery, Stanford University, Stanford, California, USA (M.Z., T.J.W.); Department of Radiology, Stanford University, Stanford, California, USA (M.Z., E.T., S.W., S.B., J.S., S.N., K.W.Y.); Department of Neurosurgery, University of Utah, Salt Lake City, Utah, USA (F.H., M.A.M.); Department of Radiology, Tehran University of Medical Sciences, Tehran, Iran (M.M.); Stanford School of Medicine, Stanford University, Stanford, California, USA (V.R., N.F.H.); Department of Neurosurgery, Mayo Clinic, Rochester, Minnesota, USA (C.P., B.W.S., R.J.S.)

Corresponding Author: Thomas J. Wilson, MD, MPH, Department of Neurosurgery, Stanford University, 300 Pasteur Drive, R293, Stanford, CA 94305, USA (wilsontj@stanford.edu).

[†]These authors contributed equally to this work.

Abstract

Background. Non-invasive differentiation between schwannomas and neurofibromas is important for appropriate management, preoperative counseling, and surgical planning, but has proven difficult using conventional imaging. The objective of this study was to develop and evaluate machine learning approaches for differentiating peripheral schwannomas from neurofibromas.

Methods. We assembled a cohort of schwannomas and neurofibromas from 3 independent institutions and extracted high-dimensional radiomic features from gadolinium-enhanced, T1-weighted MRI using the PyRadiomics package on Quantitative Imaging Feature Pipeline. Age, sex, neurogenetic syndrome, spontaneous pain, and motor deficit were recorded. We evaluated the performance of 6 radiomics-based classifier models with and without clinical features and compared model performance against human expert evaluators.

Results. One hundred and seven schwannomas and 59 neurofibromas were included. The primary models included both clinical and imaging data. The accuracy of the human evaluators (0.765) did not significantly exceed the no-information rate (NIR), whereas the Support Vector Machine (0.929), Logistic Regression (0.929), and Random Forest (0.905) classifiers exceeded the NIR. Using the method of DeLong, the AUCs for the Logistic Regression (AUC = 0.923) and K Nearest Neighbor (AUC = 0.923) classifiers were significantly greater than the human evaluators (AUC = 0.766; $p = 0.041$).

Conclusions. The radiomics-based classifiers developed here proved to be more accurate and had a higher AUC on the ROC curve than expert human evaluators. This demonstrates that radiomics using routine MRI sequences and clinical features can aid in differentiation of peripheral schwannomas and neurofibromas.

Key Points

1. Differentiation between schwannomas and neurofibromas is difficult using conventional imaging.
2. Our machine learning algorithm is highly accurate in diagnosing benign nerve tumors.
3. Radiomics using conventional MRI sequences outperforms expert humans in diagnosing BPNSTs.

Importance of the Study

Non-invasive differentiation between schwannomas and neurofibromas is important for appropriate management, preoperative counseling, and surgical planning but is difficult using conventional imaging. Radiomics-based machine learning classifiers incorporating conventional MRI and clinical data outperformed expert human evaluators for accuracy and area under the curve (AUC) for the receiver operating characteristic curve. The accuracy of the human evaluators (0.765) did not significantly exceed the no-information rate (NIR), whereas the Support Vector Machine (0.929), Logistic Regression (0.929), and Random Forest (0.905) classifiers exceeded the NIR. Using the method of DeLong, the AUCs for the Logistic Regression (AUC = 0.923) and K Nearest Neighbor (AUC = 0.923) classifiers were significantly greater than the human evaluators (AUC=0.766; $p = 0.041$). This demonstrates that radiomics using routine MRI sequences and clinical features can aid in differentiation of peripheral schwannomas and neurofibromas. Improved diagnosis should allow for better management, preoperative counseling, and surgical planning for patients with peripheral nerve sheath tumors.

The most common histologic subtypes of benign peripheral nerve sheath tumors (BPNSTs) are schwannomas and neurofibromas. Representing 10–12% of benign soft tissue neoplasms, these tumors can occur sporadically or as part of a genetic syndrome, such as neurofibromatosis type 1 (NF1), NF2, and schwannomatosis.[1] While both benign, these tumors have different cells of origin. Neurofibromas are thought to originate from nonmyelinating Schwann cells and often incorporate a variety of non-neoplastic nerve elements within the tumor, including axons, perineurial cells, and fibroblasts. This results in neurofibromas often involving many fascicles within the nerve of origin.[2–4] By contrast, schwannomas are a homogeneous proliferation of myelinating Schwann cells, with a collagenous capsule as a consistent finding. Schwannomas typically involve a single fascicle in the nerve of origin, with the encapsulation limiting involvement of neighboring fascicles.[4]

Due to these observed differences in growth pattern and fascicular involvement, resection of schwannomas versus neurofibromas may have different risk profiles and likelihoods of gross total resection.[5–7] Furthermore, neurofibromas comparatively have an increased risk of malignant transformation. While differentiating between schwannomas and neurofibromas has value, biopsy of suspected benign nerve tumors carries meaningful risks.[6] Thus, non-invasive methods of differentiation would have clinical value. Despite this, differentiation has proven difficult with qualitative interpretation of conventional imaging, making there a need for improved non-invasive methods of differentiation.[8]

Radiomics enables a quantitative approach for assessing the voxel composition of a target lesion and has had early success with distinguishing amongst neurogenic, soft tissue tumors.[9–11] Such analysis has yet to be performed for schwannomas and neurofibromas. The objectives of the current study were to develop a machine learning radiomic classifier for differentiating schwannomas and neurofibromas based on conventional magnetic resonance imaging (MRI) and basic clinical data and to test this classifier for accuracy.

Methods

Study Population

Consecutive patients with a pathologic diagnosis of a peripheral (extradural) schwannoma or neurofibroma were identified at 3 participating institutions by query of the pathology databases. The institutional review boards at all 3 participating institutions approved the study, with waiver of consent. The workflow is shown in Supplementary Figure 1. Patients were included if they had both a high-quality preoperative MRI and a tumor specimen to serve as ground truth for pathology. Patients were excluded if the MRI had significant motion degradation or if the MRI was non-diagnostic due to other artifacts or poor quality. Among the 211 patients with BPNSTs (Stanford University, $n = 54$; University of Utah, $n = 131$; Mayo Clinic, $n = 26$) submitted for quality control, axial, T1-weighted, gadolinium-enhanced (T1-gad) imaging was identified as the most commonly acquired imaging across the participating centers. Due to an insufficient number of proton density (PD) and T2-weighted (T2W) imaging available in a consistent imaging plane (sagittal, coronal, or axial), only axial T1-gad images were used for computational image feature extraction and radiomic classifier development.

Clinical Variables

For each identified patient, chart review was utilized to abstract demographic and clinical variables. Variables that were abstracted for analysis included age at the time of the operation, sex, neurogenetic diagnosis (NF1, NF2, or schwannomatosis), presence of spontaneous pain, and presence of a preoperative motor deficit.

Image Segmentation, Pre-processing, and Feature Extraction

ITK-SNAP (University of Pennsylvania) was used by 2 board-certified neuroradiologists (ET, KY) to

delineate the volumetric regions of interest for each BPNST. The PyRadiomics package on the Quantitative Imaging Feature Pipeline (QIFP) was then used to automatically extract 900 image features, including 1st order statistics, 2D/3D Shape, Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Neighboring Gray Tone Difference Matrix (NGTDM), and Gray Level Dependence Matrix (GLDM), as defined by the Imaging Biomarker Standardization Initiative.[12,13] The 1st order features assess individual voxels, without concern for spatial relationships with other voxels, while matrix features assess voxel relationships. MRI studies were normalized for voxel size (1 x 1 x 1 mm) and intensity (scale factor of 100). A fixed bin width of 10 was used for gray value discretization. Pre-processing filters included wavelet (8 coefficients) and Laplacian of Gaussian (3 sigma). Feature extraction was calculated for classes, including 1st order statistics, shape descriptors, and gray-level derivatives.[14]

Feature Selection and Validation

Training and test sets were randomly allocated from the total cohort in a 75:25 ratio. For all analyses, schwannoma was designated the positive class. Sparse regression analysis by a least absolute shrinkage and selection operator (LASSO) was used for feature reduction from the training set. This was performed with 10-fold cross-validation and repeated for 1000 cycles. The mean squared error was calculated for 100 lambdas in each cycle or until a minimum was achieved. The optimal lambda was identified as the lowest mean squared error value and used for feature reduction and coefficient calculations. Both radiologic and clinical variables were incorporated at this stage into the primary model. Features that were represented in ≥ 80% of the cycles were retained for subsequent classifier optimization. Feature reduction was performed using RStudio version 1.2.5033 (RStudio, PBC).

Classifier Model Building and Analysis

The retained features were submitted to 6 training models, including support vector machine (SVM), logistic regression (LR), k-nearest neighbor (KNN), random forest (RF), extreme gradient boosting (XGB), and neural net (NN). The cohort underwent resampling to correct for sample imbalance. Optimal classifier parameters were identified by grid search (Supplementary Table 1). The final radiomics classifier was guided by maximizing the area under the curve (AUC). Confidence intervals for each metric were obtained by bootstrapping of the test sets for 2000 random samples. Relative influences of the clinical and radiologic features were calculated for logistic regression based on coefficients used in the weighted sum. Relative influence was obtained for tree-based models by averaging the number of times a variable was selected during machine-learning algorithms across iterative regression trees and normalized so that all variables summed to 100%.[15] Model training was performed using Python version 3.8.5.

The same process was repeated to generate 2 separate secondary models, 1 with the imaging features only (i.e., excluding the clinical features) and 1 with the clinical features only (i.e., excluding the imaging features).

Test Set Evaluation by Human Evaluators

For comparison with the radiomic classifiers, human evaluators were provided the test cohort of T1-weighted, post-gadolinium images plus additional T2-weighted or proton-density weighted images, when available, and then asked to classify the tumor as a schwannoma or neurofibroma. The evaluators were then provided the clinical variables associated with the images and again asked to classify the tumor as schwannoma or neurofibroma, using both the clinical variables and MRI. The diagnoses of the human evaluators were then compared against the ground truth pathologic diagnoses. The human evaluators included 2 attending peripheral nerve surgeons and 2 attending radiologists.

Expert Score

In order to compare against the classifier, an Expert Score was generated for each tumor (1 point for each individual human expert calling the tumor a schwannoma and 0 for neurofibroma). With 4 human experts evaluating each tumor, the maximum score was 4 and the minimum score was 0 (e.g., if all 4 experts diagnosed the tumor as a schwannoma, that tumor would receive an Expert Score of 4, versus if 2 experts diagnosed the tumor as a schwannoma and 2 experts diagnosed the tumor as a neurofibroma, that tumor would receive an Expert Score of 2). A receiver operating characteristic (ROC) curve was generated for the Expert Score and the optimized threshold chosen to maximize AUC.

Statistical Analysis

Categorical variables were analyzed using the χ^2 test or Fisher exact test, as appropriate, and continuous variables were analyzed using Student's *t*-test. A p-value < 0.05 was considered statistically significant for all analyses. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1 score, AUC for the ROC curve, and accuracy were calculated for each of the classifiers, the human evaluators individually, and the Expert Score. For these calculations, schwannoma was designated as "positive". The accuracy was compared against the no-information rate (NIR), the proportion of the largest class in the set. ROC curves were compared against one another using the method of DeLong.[16] All data analyses were performed using StataSE.

Results

Comparison of Clinical Variables

A total of 166 patients were included: 107 (64%) schwannomas and 59 (36%) neurofibromas (Table 1).

Table 1 Comparison of Clinical Variables Between Patients With Neurofibromas and Schwannomas

	Neurofibroma (*N* = 59)	Schwannoma (*N* = 107)	*P*-value
Age, years (SD)	39.9 (13.9)	49.7 (14.4)	<0.001
Sex			0.106
Male	38 (64%)	55 (51%)	
Female	21 (36%)	52 (49%)	
Spontaneous Pain	14 (24%)	27 (25%)	0.830
Motor Deficit	13 (22%)	30 (28%)	0.398
NF1	34 (58%)	4 (4%)	<0.001
NF2	0 (0%)	10 (9%)	0.015
Schwannomatosis	0 (0%)	5 (5%)	0.092

Abbreviation. SD, standard deviation.

Patients with neurofibromas were younger ($p < 0.001$) and more likely to have a diagnosis of NF1 ($p < 0.001$), while patients with schwannomas were more likely to carry a diagnosis of NF2 ($p = 0.015$) (Table 1). There was no significant difference in sex, history of pain, or presence of motor deficit.

Primary Model: Imaging and Clinical Features

The primary radiomics signature was identified by LASSO feature reduction using both imaging and clinical features. Following feature reduction, 3 imaging features and 1 clinical feature were consistently selected by LASSO in over 80% of regressions (Supplementary Table 2, Supplementary Figure 2). The radiomic components included 1 shape (Spherical Disproportion) and two 1st order features (Kurtosis with a Laplacian of Gaussian, sigma 1 mm filter, and Kurtosis with a Wavelet, High-Low-Low pass filter). The clinical feature included was NF1 diagnosis. These 4 selected features were used to subsequently train and validate candidate classifier models.

Among the 6 classifier models evaluated, Logistic Regression and K Nearest Neighbor showed the highest AUC (0.923), while Support Vector Machine and Logistic Regression showed the highest accuracy (0.929) and F1 score (0.955) (Figure 1, Table 2). Only Support Vector Machine, Logistic Regression, and Random Forest showed accuracy that significantly exceeded the no-information rate (Table 2). The no-information rate was 0.643, the proportion of schwannomas in the test set.

Relative influence of model variables was assessed in the Logistic Regression, Random Forest, and eXtreme Gradient Boost classifiers (Supplementary Figure 3). NF1 status consistently had the strongest contribution across classifiers. Spherical Disproportion was consistently the most influential imaging feature.

Secondary Models: Imaging Features Alone and Clinical Features Alone

A secondary model was constructed using only imaging features, excluding clinical features. Ten features were retained, and all contributed to the final classifier. These again included Kurtosis and Spherical Disproportion elements. The AUC for Neural Network on the test set was the highest at 0.897 (Supplementary Figure 4). The F1 score was highest for Logistic Regression and Neural Network models (0.814) (Supplementary Table 4). None of the classifiers had an accuracy that significantly exceeded the NIR. Spherical Disproportion had the greatest relative contribution (Supplementary Figure 5).

Another secondary model was constructed using the clinical features alone, excluding the imaging features (Supplementary Figure 6). The highest AUC for the test set was observed for the Neural Network classifier (0.877) (Supplementary Table 4). The highest accuracy was observed for the Logistic Regression classifier, with the Logistic Regression, Support Vector Machine, and K Nearest Neighbor classifiers outperforming the NIR. NF1 status had the highest contribution for the Logistic Regression and eXtreme Gradient Boost classifiers, whereas age had the highest relative importance for the Random Forest classifier (Supplementary Figure 7).

Human Evaluation of Test Set

Using both clinical and imaging data, the best human reviewer had an accuracy of 0.725, with an average accuracy of 0.681, and none of the 4 human evaluators exceeded the NIR (Supplementary Table 5). Using imaging alone, the best human evaluator had an accuracy of 0.706, with an average accuracy of 0.652, and none of the 4 evaluators significantly exceeded the NIR (Supplementary Table 5).

Expert Score Classifier

The Expert Score classifier using imaging and clinical data had an accuracy of 0.765, which did not significantly exceed the NIR. The AUC for the ROC curve was 0.766 (Figure 2, Table 3). When using imaging data alone, the Expert Score classifier achieved an accuracy of 0.758, which did not significantly exceed the no information rate. The AUC for the ROC curve was 0.734 (Supplementary Figure 8, Table 3).

Comparison of ROC Curves

The method of DeLong was used to compare AUCs. The AUC for the Logistic Regression and K Nearest Neighbor classifiers significantly exceeded the AUC for the Expert Score (Figure 2, Supplementary Table 6).

Discussion

Schwannomas and neurofibromas are difficult to differentiate due to their similar imaging appearance and clinical presentations.[17] While specific imaging findings have proven highly sensitive and specific for a less common form of BPNST, intraneural perineurioma, imaging findings that are highly sensitive and specific for schwannoma versus neurofibroma have not been defined.[18] We

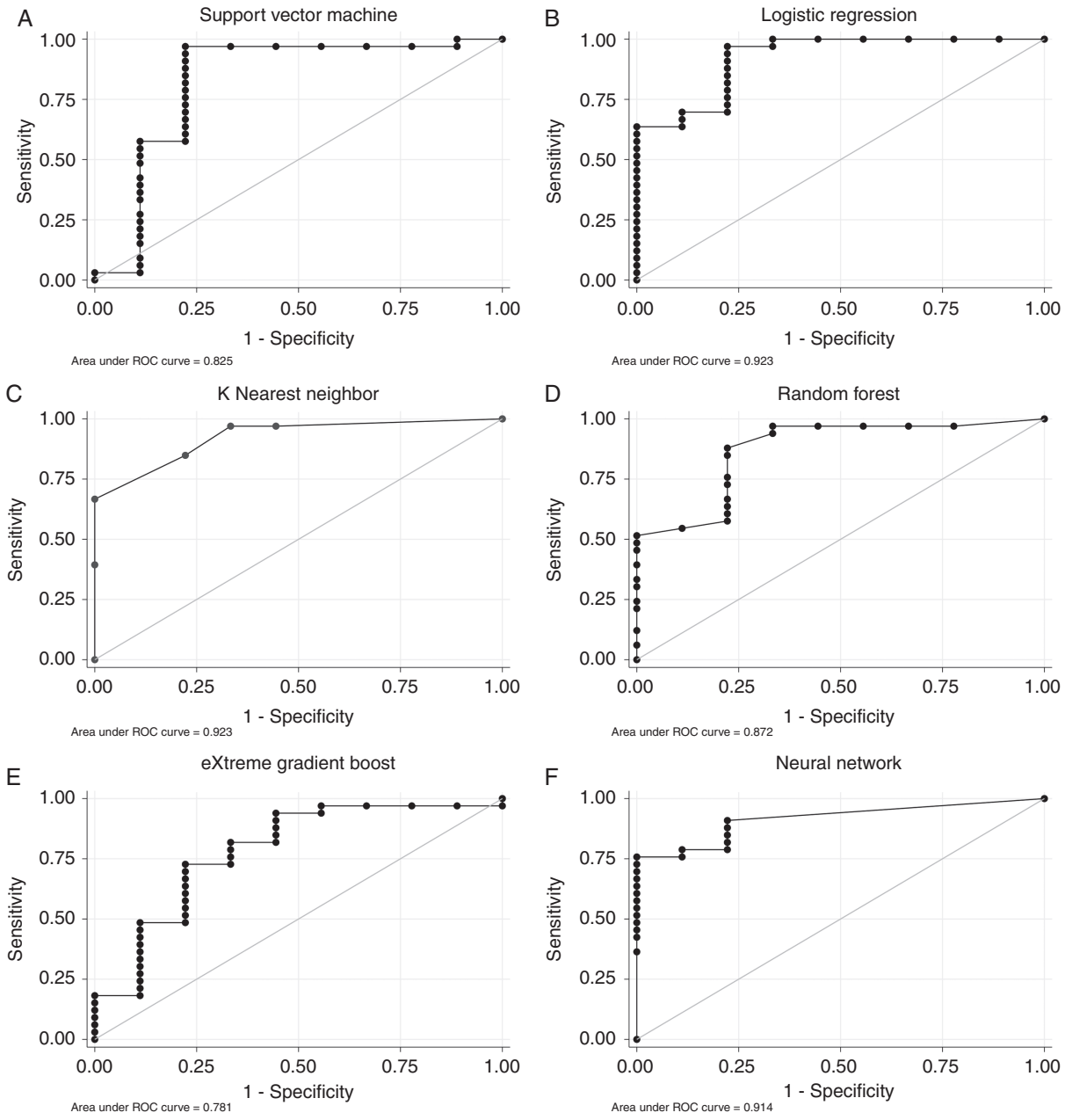

Fig. 1 Receiver operating characteristic curves measuring the performance of A. Support Vector Machine, B. Logistic Regression, C. K Nearest Neighbor, D. Random Forest, E. eXtreme Gradient Boosting, and F. Neural Network classifiers incorporating both imaging and clinical features.

hypothesized that radiomics may provide a means for differentiation using only conventional imaging and basic clinical data. We found that a radiomics approach that incorporates both imaging and clinical features outperformed human evaluators, even when human evaluators were provided more data in the form of additional MR imaging sequences. As a first iteration, this represents a significant advance.

Why does differentiation between schwannomas and neurofibromas matter? Schwannomas and neurofibromas appear to have different risk profiles during resection and different likelihoods of gross total resection. Schwannomas are typically well-encapsulated tumors that involve a single fascicle of the nerve. By defining the plane between the capsule and the layers of pseudocapsule, schwannomas can be separated from the uninvolved fascicles of the nerve, allowing for function-sparing complete resection in many cases.[19–21] Conversely, neurofibromas do not have a well-defined capsule and often involve many or even most of the fascicles of the nerve. This theoretically makes function-sparing complete resection less likely, and some studies have shown that resection of neurofibromas

Table 2 Performance Metrics by Six Classifiers When Provided Clinical and Imaging Features

	SVM	LR	KNN	RF	XGB	NN
AUC (95% CI)	0.825 (0.567–1.000)	0.923 (0.798–1.000)	0.923 (0.828–0.988)	0.872 (0.719–0.987)	0.781 (0.569–0.951)	0.914 (0.832–0.996)
Accuracy	0.929*	0.929*	0.833	0.905*	0.786	0.810
Specificity	0.778	0.778	0.778	0.667	0.667	0.889
Sensitivity	0.970	0.970	0.849	0.970	0.818	0.788
F1 Score	0.955	0.955	0.889	0.941	0.857	0.867
PPV	0.941	0.941	0.933	0.914	0.900	0.963
NPV	0.875	0.875	0.583	0.857	0.500	0.533

Abbreviations. AUC, area under the ROC curve; 95% CI, 95% confidence interval; KNN, k nearest neighbor; LR, logistic regression; NN, neural net; NPV, negative predictive value; PPV, positive predictive value; RF, random forest; SVM, support vector machine; XGB, eXtreme gradient boosting.
*Accuracy measured statistically greater than the No Information Rate (case prevalence) with *P*-value < 0.05.

is associated with higher risk of a postoperative neurological deficit and/or a lower likelihood of gross total resection.[5,22] Furthermore, neurofibromas have higher risk of malignant transformation compared to schwannomas. Thus, reliable differentiation between schwannomas and neurofibromas would allow for a more appropriate discussion with patients regarding the risks and benefits of resection and may alter the long-term management strategy. If differentiation is important, this begs the question, why not biopsy these masses to establish a tissue diagnosis? Biopsy of suspected benign nerve sheath tumors has been shown to have significant risks, including creating a neurologic deficit directly related to the biopsy and increased risk of a neurologic deficit after resection.[6] Based on this, biopsy of suspected benign nerve sheath tumors is not advisable, which serves to emphasize the significance of a technique that allows non-invasive differentiation between schwannomas and neurofibromas.

Our findings confirm the poor ability of human evaluators, including experts in the field, to differentiate between schwannomas and neurofibromas using clinical data and conventional MRI. The best performing human expert had an accuracy of 0.725, which did not significantly exceed the no-information rate. When 4 experts were combined into an Expert Score, the human experts fared no better. The accuracy was 0.765, again not exceeding the no-information rate, and the AUC for the ROC curve was 0.766. When considering both accuracy and AUC, the best performing radiomics-based classifier was the Logistic Regression classifier using both imaging and clinical features. The accuracy (0.929) significantly exceeded the no-information rate, and the AUC (0.923) was significantly higher than the AUC for the human Expert Score. The superior performance of the radiomics-based classifier was observed despite the human evaluators being given both T1-gad and T2-weighted images to evaluate, whereas T1-gad was the only imaging sequence incorporated into the machine learning classifiers. Our primary radiomics model provides evidence that quantitative, voxel-based evaluation of segmented lesions can improve diagnosis.

Using both clinical and imaging features produced a parsimonious radiomics signature with only 4 variables. We identified that Spherical Disproportion and Kurtosis were consistently preserved across validation trials. These imaging features can plausibly be linked to previously described radiographic patterns. Spherical Disproportion describes the ratio between the surface area of segmented tumor and the surface area of a sphere with a volume equal to that of the segmented tumor. Thus, it is inversely correlated with how spherical a lesion measures. Schwannomas are often unilobular, well-encapsulated, and eccentrically positioned within the associated nerve, assuming a more spherical shape compared to neurofibromas.[23–25]

Kurtosis is an imaging feature that emphasizes the distribution of voxel intensities within the segmented region. A higher value signifies a concentration of signal towards the center and more rapid decay of signal intensity from the peak. Both neurofibromas and schwannomas exhibit leptokurtic distributions, which implies distributions greater than a standard normal distribution. However, the higher mean value for neurofibromas suggests greater likelihood of a strong focus of signal. Jee and colleagues have previously described a higher likelihood of central enhancement, as opposed to diffuse or heterogeneous enhancement, compared to schwannomas.[26] This would match the finding of Kurtosis as a predictor, since strong central enhancement would have a higher Kurtosis value compared to heterogeneous or diffuse enhancement. It is important to remember that the radiomics classifier only assessed T1-weighted, post-gadolinium images, so some of the previously described findings that are T2-weighted imaging-based, such as the target sign, could not be quantitatively assessed in this iteration.[27]

The strongest predictor in the model that included imaging and clinical features was NF1 status. This finding helps internally validate the primary model. NF1 status should be a strong predictor of histologic subtype, since neurofibroma is the dominant histologic subtype in patients with NF1. Interestingly, NF2 and schwannomatosis status were not preserved by LASSO regression in the model for both imaging and clinical features. This is likely due to the fact that the data set contained a small number of NF2-positive and schwannomatosis-positive patients. We would predict that with a larger data set that contains a higher number of patients with NF2 and/or

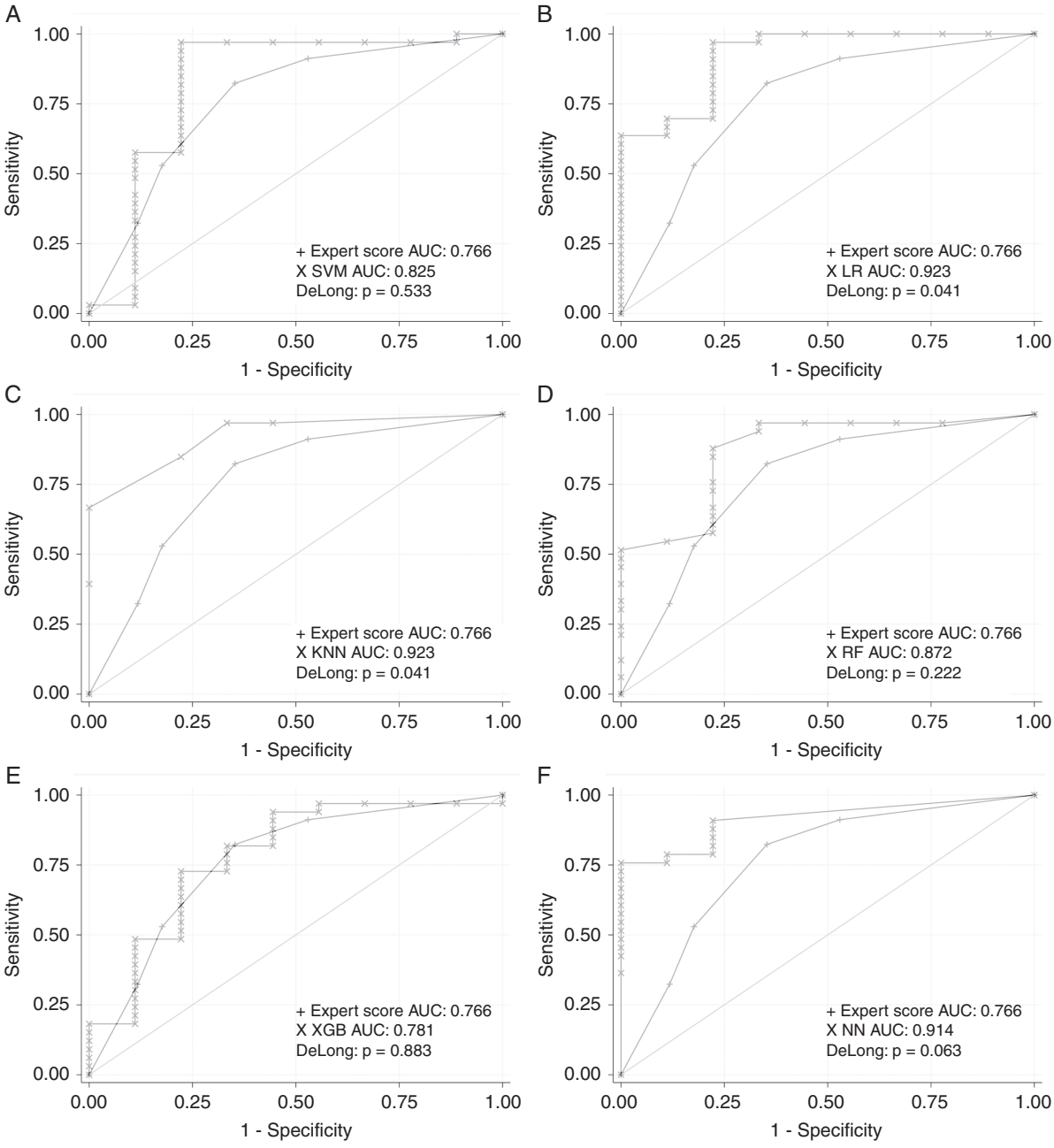

Fig. 2 Receiver operating characteristic curves comparing the human Expert Score to A. Support Vector Machine, B. Logistic Regression, C. K Nearest Neighbor, D. Random Forest, E. eXtreme Gradient Boosting, and F. Neural Network classifiers incorporating both imaging and clinical features.

Table 3 Metrics Achieved by the Expert Score Classifier

	Sensitivity	Specificity	PPV	NPV	F1 Score	Accuracy	AUC (95% CI)
Imaging + Clinical	0.848	0.611	0.800	0.688	0.824	0.765	0.766 (0.620–0.913)
Imaging Only	0.758	0.611	0.781	0.579	0.769	0.758	0.734 (0.580–0.887)

The optimal threshold was found to be ≥ 2 to diagnose schwannoma (positive class). Abbreviations. PPV, Positive Predictive Value; NPV, Negative Predictive Value; AUC, Area Under Curve; 95% CI, 95% Confidence Interval.

schwannomatosis, these parameters would become important predictors in the model.

The AUC was highest for the primary model incorporating imaging and clinical data, higher than models incorporating imaging alone or clinical data alone. The sample size did not allow for a subgroup analysis comparing syndromic (NF1, NF2, schwannomatosis) to non-syndromic tumors. We plan to do such an analysis in the future with a larger sample size. A diagnosis of NF1, NF2, or schwannomatosis is very strongly associated with tumor type (neurofibroma for NF1 versus schwannoma for NF2 or schwannomatosis), such that we would predict that in the syndromic tumors imaging does not add significantly to the clinical data. Conversely, we would hypothesize that imaging features would dominate predictive models evaluating only non-syndromic tumors. In the current study that includes syndromic and non-syndromic tumors, following this line of thinking, we would predict that both clinical and imaging features would be important, and this is likely why we observed the highest AUC for the primary model incorporating both.

Limitations

Our classification algorithm shares many of the common challenges in radiomics that limits its performance. These include heterogeneity in MR acquisition technique, such as institutional protocols, machine technology, and image quality. Heterogeneity in patient/tumor parameters, such as anatomic location, may similarly attenuate the predictive power of the algorithm. We were limited in our ability to include T2-weighted sequences in this iteration of the classifier. There is wide variance in the specific T2-weighted sequence and acquisition plane (i.e., axial, sagittal, coronal) from institution to institution and body region to body region. The heterogeneous T2-weighted imaging acquisition meant that we could not include T2-weighted imaging in this iteration of the classifier.

The sample size may have limited our discriminatory power to identify important features. Analysis of a larger sample in the future may uncover features not included in the current model that are important contributors to differentiating neurofibromas from schwannomas. The sample size also prevented us from using deep machine learning techniques that could analyze the entire image rather than just the segmented tumor volume.

There may have been bias introduced by the requirement to have a tissue diagnosis. Tumors that are imaged and operated may be a different population of tumors than those that are asymptomatic and for which a tissue diagnosis is never obtained. Thus, a conservative approach would be to apply this classifier and to extrapolate the outcomes to only those tumors for which surgery is being considered.

Supplementary material

Supplementary material is available at *Neuro-Oncology* online.

Keywords

machine learning | nerve sheath tumor | neurofibroma | radiomics | schwannoma

Funding

This study was funded in part by the Medtronic Young Clinician Investigator Award from the Neurosurgery Research & Education Foundation (NREF). Michael Zhang is funded by a T32 award from the National Institutes of Health [5T32CA009695-27 (MPI)]. Sandy Napel is funded by a National Institutes of Health U01 grant [CA187947].

Conflict of interest statement. The authors have no conflicts of interest to declare.

Authorship statement. Conception and design, acquisition of data, or analysis and interpretation of data: M.Z., E.T., S.W., F.H., M.M., V.R., C.P., B.W.S., N.F.H., S.B., J.S., S.N., R.J.S., M.A.M., K.W.Y., and T.J.W. Drafting the article or critical revision: M.Z., R.J.S., M.A.M., K.W.Y., and T.J.W. Final approval of the manuscript: M.Z., E.T., S.W., F.H., M.M., V.R., C.P., B.W.S., N.F.H., S.B., J.S., S.N., R.J.S., M.A.M., K.W.Y., and T.J.W.

References

1. Kransdorf MJ. Benign soft-tissue tumors in a large referral population: distribution of specific diagnoses by age, sex, and location. *AJR Am J Roentgenol.* 1995;164(2):395–402.
2. Weiss SW, Nickoloff BJ. CD-34 is expressed by a distinctive cell population in peripheral nerve, nerve sheath tumors, and related lesions. *Am J Surg Pathol.* 1993;17(10):1039–1045.
3. Carroll SL. Molecular mechanisms promoting the pathogenesis of Schwann cell neoplasms. *Acta Neuropathol.* 2012;123(3):321–348.
4. Rodriguez FJ, Folpe AL, Giannini C, Perry A. Pathology of peripheral nerve sheath tumors: diagnostic overview and update on selected diagnostic problems. *Acta Neuropathol.* 2012;123(3):295–319.
5. Donner TR, Voorhies RM, Kline DG. Neural sheath tumors of major nerves. *J Neurosurg.* 1994;81(3):362–373.
6. Perez-Roman RJ, Shelby Burks S, Debs L, Cajigas I, Levi AD. The risk of peripheral nerve tumor biopsy in suspected benign etiologies. *Neurosurgery.* 2020;86(3):E326–E332.
7. Wilson TJ, Hamrick F, Alzahrani S, et al. Analysis of the effect of intraoperative neuromonitoring during resection of benign nerve sheath tumors on gross total resection and neurological complications. *J Neurosurg.* 2021:1–10.
8. Karsy M, Guan J, Ravindra VM, Stilwill S, Mahan MA. Diagnostic quality of magnetic resonance imaging interpretation for peripheral nerve sheath tumors: can malignancy be determined? *J Neurol Surg A Cent Eur Neurosurg.* 2016;77(6):495–504.

9. Ghosh A, Malla SR, Bhalla AS, Manchanda S, Kandasamy D, Kumar R. Texture analysis of routine T2 weighted fat-saturated images can identify head and neck paragangliomas - A pilot study. *Eur J Radiol Open.* 2020;7:100248.
10. Uthoff J, De Stefano FA, Panzer K, et al. Radiomic biomarkers informative of cancerous transformation in neurofibromatosis-1 plexiform tumors. *J Neuroradiol.* 2019;46(3):179–185.
11. Zhang M, Tong E, Hamrick F, et al. Machine-learning approacch to diffferentiation of benign and malignant peripheral nerve sheath tumors: a multi-center study. *Neurosurgery.* 2021;89(3):509–517.
12. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 2017;77(21):e104–e107.
13. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology.* 2020;295(2):328–338.
14. Echegaray S, Bakr S, Rubin DL, Napel S. Quantitative Image Feature Engine (QIFE): an open-source, modular engine for 3D quantitative feature extraction from volumetric medical images. *J Digit Imaging.* 2018;31(4):403–414.
15. Friedman JH, Meulman JJ. Multiple additive regression trees with application in epidemiology. *Stat Med.* 2003;22(9):1365–1381.
16. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44(3):837–845.
17. Chee DW, Peh WC, Shek TW. Pictorial essay: imaging of peripheral nerve sheath tumours. *Can Assoc Radiol J.* 2011;62(3):176–182.
18. Wilson TJ, Howe BM, Stewart SA, Spinner RJ, Amrami KK. Clinicoradiological features of intraneural perineuriomas obviate the need for tissue diagnosis. *J Neurosurg.* 2018;129(4):1034–1040.
19. Stone JJ, Boland JM, Spinner RJ. Analysis of peripheral nerve schwannoma pseudocapsule. *World Neurosurg.* 2018;119:e986–e990.
20. Stone JJ, Puffer RC, Spinner RJ. Interfascicular resection of benign peripheral nerve sheath tumors. *JBJS Essent Surg Tech.* 2019;9(2):e18.
21. Stone JJ, Spinner RJ. Go for the Gold: a "Plane" and simple technique for resecting benign peripheral nerve sheath tumors. *Oper Neurosurg (Hagerstown).* 2020;18(1):60–68.
22. Wilson TJ, Hamrick F, Alzahrani S, et al. Analysis of the effect of intraoperative neuromonitoring during resection of benign nerve sheath tumors on gross-total resection and neurological complications. *J Neurosurg.* 2021:1–10.
23. Antonescu CR, Perry A, Woodruff JM. Schwannoma (including variants). In: Fletcher CDM, Bridge JA, Hogendoorn PCW, Mertens F, eds. *World Health Organization Classification of Tumours of Soft Tissue and Bone.* Vol. 5. Lyon: IARC Press; 2013:170–172.
24. Cerofolini E, Landi A, DeSantis G, Maiorana A, Canossi G, Romagnoli R. MR of benign peripheral nerve sheath tumors. *J Comput Assist Tomogr.* 1991;15(4):593–597.
25. Pilavaki M, Chourmouzi D, Kiziridou A, Skordalaki A, Zarampoukas T, Drevelengas A. Imaging of peripheral nerve sheath tumors with pathologic correlation: pictorial review. *Eur J Radiol.* 2004;52(3):229–239.
26. Jee WH, Oh SN, McCauley T, et al. Extraaxial neurofibromas versus neurilemmomas: discrimination with MRI. *AJR Am J Roentgenol.* 2004;183(3):629–633.
27. Kakkar C, Shetty CM, Koteshwara P, Bajpai S. Telltale signs of peripheral neurogenic tumors on magnetic resonance imaging. *Indian J Radiol Imaging.* 2015;25(4):453–458.