Deep Learning-based Detection of Malignant Signature in Skull Base Schwannomas: Bicentric Study and Prognostic Modeling

Daniela Stastna,MD[1,2], Danielle Dang,MD[1], Richard Mannion,MD[2], Patrick Axon[2], James Tysome[2], Neil Donnelly[2], Danielle Borsetto[2], Carolina Giannini[1], Rushna Ali[1], Ari Ercole[2], Jonathan Coles[2], Robert Spinner[1], Robert Macfarlane[2], Jamie Van Gompel[1] & Michael Link[1]

[1]Neurosurgery Dpt, Mayo Clinic, RST, USA ; [1]University of Cambridge, UK,

Abstract

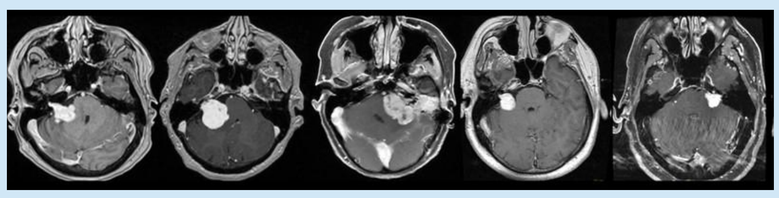
In this bicentric retrospective study, we developed a machine-learning framework to detect malignant MRI signatures of cranial nerve malignant peripheral nerve sheath tumors (CN-MPNSTs). Ten CN-MPNSTs were compared with 850 benign schwannomas, with augmentation. Deep-learning and radiomics features from preoperative CE-T1wMRI were used to train XGBoost and a Multilayer Perceptron (MLP), achieving excellent discrimination (AUC 0.93–0.99). Gross-total resection improved PFS, and a SHAP-derived RadScore independently stratified prognosis.

Introduction

Malignant peripheral nerve sheath tumors of cranial nerves (CN-MPNSTs) are rare, aggressive tumors. Malignant transformation occurs in <1% of benign CN schwannomas. Prognosis remains poor, with 5-year survival of 30–60%. Timely identification of CN-MPNSTs is critical, yet routine MRI lacks this capability.

We aimed to (1) develop a machine learning (ML) framework for detecting the malignant signature of CN-MPNSTs on baseline MRI, and (2) identify clinico-radiological predictors of prognosis.

1a



1b.

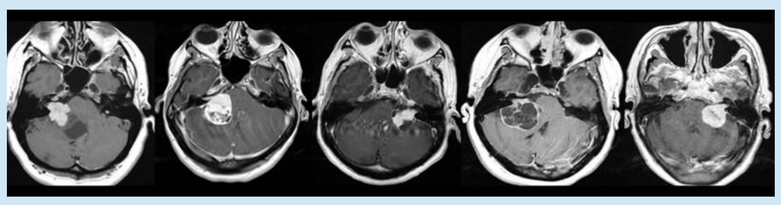


Figure 1:

Axial CET1wMRI of CN-MPNSTs (1a) in comparison with (1b) benign schwannomas

All depicted are vestibular schwannomas.

Methods and Materials

This bicentric retrospective study included 10 patients with resected cranial nerve malignant peripheral nerve sheath tumors (CN-MPNSTs). Models were trained against 850 sporadic benign cranial nerve schwannomas (840 vestibular, 10 trigeminal). To mitigate extreme class imbalance, 20 peripheral nerve MPNST fragments were incorporated into hybrid-synthetic malignant cases, and additional oversampling strategies (ADASYN-like, Borderline-SMOTE) were applied.

Preoperative contrast-enhanced T1-weighted MRI (CE-T1wMRI) underwent registration, isotropic resampling, bias-field correction (N4), intensity normalization, and multi-scale filtering. Tumors were segmented using ITK-SNAP, yielding 1,469 radiomic and deep learning–derived features per ROI, including texture, morphology, and interpretable CNN-based heterogeneity descriptors.

Feature-based classifiers (XGBoost and 4-layer SNN/MLP) were evaluated using nested stratified cross-validation with bootstrapped 95% confidence intervals. Model interpretability was assessed using SHAP to identify stable malignant signatures. Prognostic modeling used an Elastic Net–regularized Cox model to derive a radiomic risk score (RadScore), stratifying progression risk via Kaplan–Meier analysis.

Results

Ten patients with CN-MPNSTs (median age 57 years; 60% male) were compared with 850 benign cranial nerve schwannomas (median age 64 years; p=0.022). NF2 prevalence was similar (10% vs 8%). Most CN-MPNSTs were radiation-induced (70%; latency 6.5–26 years), and the most frequently affected nerves were CN VIII (50%) and CN V (30%). Gross-total resection (GTR) was achieved in 50%, while 90% received adjuvant radiotherapy. Tumors showed markedly elevated Ki-67 (median 50%) and loss of H3K27me3 in radiation-induced cases. Median PFS and OS were 0.6 and 0.9 years, respectively. GTR significantly prolonged PFS compared with incomplete resection (2.6 vs 0.3 years; p=0.0033; HR=25.3) and predicted improved OS (p=0.0015).

XGBoost demonstrated strong discrimination but unstable minority-class precision on the original cohort (AUC 0.93; accuracy 0.94). Hybrid augmentation markedly improved sensitivity, with recall increasing to 0.90–0.94 and F1-scores to 0.87–0.90, while AUC remained high (0.98–0.99). The MLP achieved similarly robust performance across dataset variants (AUC 0.97–0.99; recall up to 0.94).

SHAP analyses showed that approximately 40 features accounted for ≥80% of model explanatory power. Dominant predictors included prototype-similarity metrics, wavelet-based texture features, shape irregularity descriptors, and DL-derived rim–core attention measures, reflecting marked intratumoral heterogeneity. An Elastic Net–derived RadScore significantly stratified patients into high- and low-risk groups (p=0.015; HR=8.3) and remained independently associated with PFS alongside extent of resection.

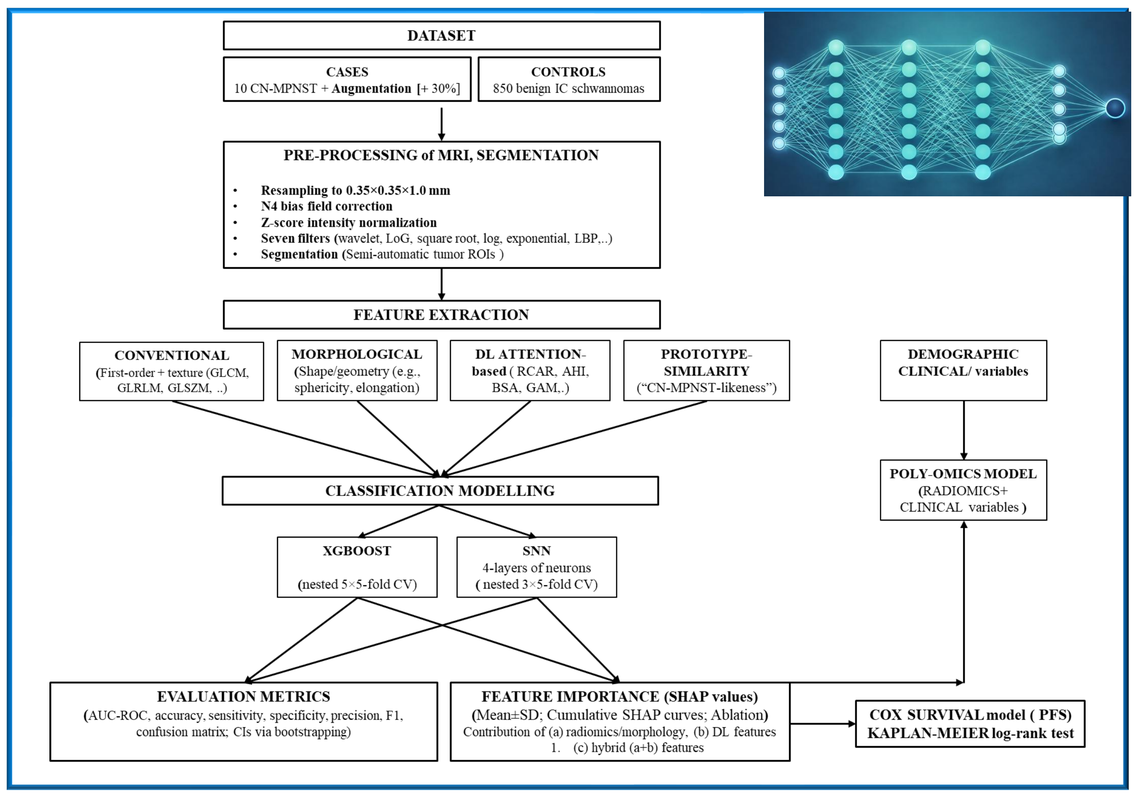


Figure 2: Pipeline of machine learning models using hybrid deep learning +radiomics features

Results

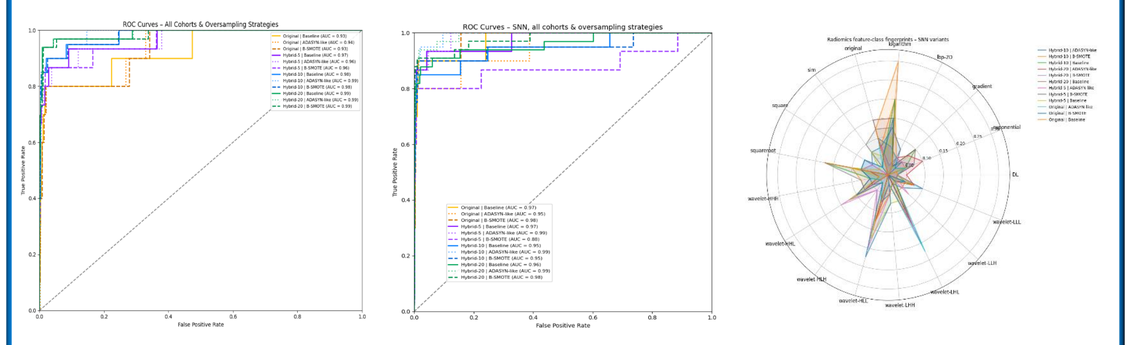


Figure 3. 3a) Performance of XGBoost and MPL/SNN models distinguishing CN-MPNSTs from benign tumors: ROC curves for 12 XGBoost training strategies. 3c) Feature importance and radiomic signature analyses for the XGBoost model with radar plot illustrating dominant feature-class fingerprints shared across all 12 XGBoost variants.

Univariate Cox analysis						
Variable	***coef***	***HR***	***HR_lower_95***	***HR_upper_95***	***z***	***p***
Extent of resection	1.482789	4.405215	1.271282	15.264843	2.338509	***0.019361***
RadScore	1.224161	3.40131	1.15293	10.034356	2.217776	***0.02657***
Ki-67	0.030079	1.030536	0.98942	1.07336	1.44793	*0.147637*
NF1 status	1.722314	5.597466	0.430024	72.860178	1.315422	*0.188368*
CN origin	0.334142	1.396741	0.791466	2.464902	1.152983	*0.248918*
Etiology (RT-induced)	0.435431	1.54563	0.32046	7.454826	0.542401	*0.587542*
Age	0.013692	1.013786	0.926153	1.109711	0.296823	*0.766602*
Gender	0.008388	1.008423	0.209859	4.845712	0.010473	*0.991644*

Table 1: Univariate Cox analysis of risk clinico-radiological variables in relation to progression-free survival.

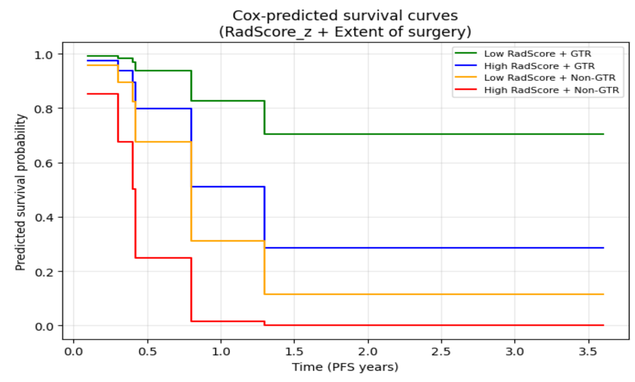


Figure 7:

Progression-free survival curves derived from penalized Cox proportional hazards model, stratified by radiomic risk score (z-scored *RadScore*) and extent of resection. Patients were dichotomized into HighRad and LowRad groups using the median RadScore. *Legend: GTR- gross-total resection.*

Discussion

In this bicentric cohort with heterogeneous MRI acquisition, a DL-radiomics framework reliably distinguished CN-MPNSTs from benign cranial schwannomas on routine CE-T1wMRI. Despite extreme class imbalance, both XGBoost and SNN/MLP achieved excellent discrimination and identified a consistent malignant signature driven by texture heterogeneity and shape irregularity. Hybrid augmentation improved sensitivity, while a SHAP-derived RadScore independently predicted progression. External validation is required.

Conclusions

Despite the rarity of CN-MPNSTs and bicentric MRI heterogeneity, augmented DL-radiomics models achieved stable malignant-signature discrimination and prognostic stratification. These findings support the feasibility of an interpretable MRI-based AI tool for rare CN-MPNSTs, while highlighting the need for multicenter validation. Early detection of malignant transformation on routine MRI may enable timelier intervention and improved outcomes.

Contact

[Daniela Stastna
Dept of Neurosurgery, Mayo clinic]
[Rochester, MN

References

1. Behling F, Bersali I, Santacroce A, et al. Transition of a vestibular schwannoma to a malignant peripheral nerve sheath tumor with loss of H3K27 trimethylation after radiosurgery-a case report and review of the literature. *Neurosurg Rev*. 2022;45(1):915-922. doi:10.1007/s10143-021-01620-3