Convolutional Neural Networks Accurately Predict Benign versus Malignant Status Among Peripheral Nerve Sheath Tumors

Alexander Mazal1, Liyuan Chen2, Feng Poh3, Jing Wang2, Michael Folkert2, Oganess Ashikyan4, Parham Pezeshk4 and Avneesh Chhabra4*

1Department of Surgery and Perioperative Care, The University of Texas at Austin Dell Medical School, Austin, TX, USA
2Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, Texas, USA
3Mount Elizabeth Hospital, Singapore
4Division of Musculoskeletal Radiology, Department of Radiology, UT Southwestern Medical Center, Dallas, TX, USA

*Correspondence to:
Avneesh Chhabra, MD
Associate Professor of Radiology
Chief, Division of Musculoskeletal Radiology
UT Southwestern Medical Center
Harry Hines Blvd, Dallas, TX, USA
E-mail: Avneesh.Chhabra@UTSouthwestern.edu

Received: January 11, 2021
Accepted: February 19, 2021
Published: February 22, 2021

Citation: Mazal A, Chen L, Poh F, Wang J, Ashikyan O, et al. 2021. Convolutional Neural Networks Accurately Predict Benign versus Malignant Status Among Peripheral Nerve Sheath Tumors. J Neuroimaging Psychiatry Neurol 6(1): 1-6.

Copyright: © 2021 Mazal et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY) (http://creativecommons.org/licenses/by/4.0/) which permits commercial use, including reproduction, adaptation, and distribution of the article provided the original author and source are credited.

Published by United Scientific Group

Abstract

Background: Peripheral nerve sheath tumors (PNSTs) comprise ~5-10% of soft tissue tumors encountered in the clinical setting. Benign lesions (BPNSTs), such as neurofibromas and schwannomas are often asymptomatic or cause neuropathy. Malignant peripheral nerve sheath tumors (MPNSTs) frequently exhibit rapid invasive behavior and metastatic spread. MR imaging markers do not reliably differentiate BPNSTs from MPNSTs. Convolutional neural networks employ machine learning and multi-order statistics to derive imaging signatures that could improve diagnostic assessment of PNSTs.

Purpose: To evaluate whether convolutional neural networks can accurately differentiate BPNSTs from MPNSTs and compare the accuracy to that of expert radiologist interpretation.

Materials and Methods: MR images from 47 patients with histologically-confirmed PNSTs were identified. Two separate convolutional neural networks (CNNs) were created using fat-suppressed T2-weighted (fsT2W) images alone (CNN 1), and fsT2W images in combination with pre- and post-contrast T1-weighted imaging (CNN 2). CNN performance was compared to interpretation by two experienced radiologists.

Results: CNN 1 performed comparably to the radiologists, achieving an accuracy and area under the curve (AUC) of 87% and 0.89, respectively. By comparison, radiologist 1 and 2 achieved accuracies and AUCs of 73%, 0.83 and 93%, 0.83, respectively. No significant differences were found between the accuracies or AUCs of either radiologist and CNN 1 (p>0.05). CNN 2 achieved an accuracy and AUC of 93%, 0.94. Using all image sequences together, radiologists 1 and 2 achieved accuracies and AUCs of 71%, 0.81 and 71%, 0.70, respectively.

Conclusion: Convolutional neural networks accurately differentiated BPNSTs versus MPNSTs in our investigation. Larger studies may be needed to validate these results.

Keywords
Radiomic analysis, Machine learning, Convolutional neural network, Peripheral nerve sheath tumor, Neurofibroma, Schwannoma, Malignant peripheral nerve sheath tumor
Introduction

PNSTs are a group of soft tissue neoplasms of Schwann cell and/or perineurial cell origin. They are broadly classified into benign PNSTs (BPNSTs) and malignant PNSTs (MPNSTs) based on their pathologic findings and clinical invasive behavior. Among BPNSTs, the most commonly observed subtypes are neurofibromas, schwannomas, and, more rarely, perineuriomas. BPNSTs may manifest clinically with neuropathy or compressive symptoms due to local growth and invasion, or they may remain asymptomatic. Neurofibromas, in particular, typically have fascicles of origin which are non-functional [1]. By comparison, MPNSTs exhibit rapid growth with early metastatic spread [2]. They are associated with significant morbidity and mortality and confer a bleak prognosis. Early identification and treatment of MPNSTs is essential for the prevention of serious sequelae [3]. MR imaging represents the gold standard for initial identification and differentiation of these neoplasms; however, there can often be considerable overlap in the imaging features of benign and malignant tumors (Figure 1). For this reason, conventional reader-based diagnostic approaches do not reliably differentiate “indeterminate” lesions, and a biopsy is required for definitive pathologic diagnosis. Due to the high prevalence of BPNSTs with borderline features concerning for malignancy, many patients with these lesions are unnecessarily referred for biopsy rather than follow-up [4]. Although adjuncts to conventional MR imaging have been purported to increase the diagnostic yield of initial imaging, including FDG18-PET imaging, diffusion-weighted imaging (DWI) and diffusion tensor imaging (DTI), further work is needed to determine the efficacy of these techniques [5]. In a 2014 study by Demehri et al., more than 1/3 of PNSTs with both conventional and functional imaging features that were reported as being highly concerning for malignancy (avg. size ≥4.2 cm and apparent diffusion coefficient (ADC) value <1.0 x 10-3 mm2/s), were in-fact benign [4]. More robust diagnostic imaging techniques are therefore currently being explored to improve diagnostic accuracy.

Radiomics is an emerging field, which represents the convergence of several disciplines, including radiology, machine learning, and computer vision [6]. Radiomic analysis relies on automated quantitative image analytics and the extraction of quantitative image features (referred to henceforth as radiomic features), which can be leveraged to improve medical decision making. Radiomic features such as shape, volume, texture (gray level co-occurrence matrix, intensity (i.e., first order statistics, or wavelet, among others) can offer insight into tumor phenotypic characteristics [7].

There has been considerable progress in the field of radiomics and convolutional neural networks (CNNs) over the past decade, with many studies demonstrating the potential utility of these techniques in tumor prognostication, staging, screening, survival, and recurrence risk [7-15]. Nevertheless, data concerning the utility of CNNs and radiomics in the domain of PNSTs remain sparse. This study evaluated the accuracy of convolutional neural networks in the differentiation of histologically proven benign and malignant PNSTs and compared it to that of expert radiologist interpretation. We hypothesized that trained CNN models would outperform the expert radiologist interpretations.

Methods

The study was conducted under institutional review board approval and the informed consent was waived.

Study population

MR images from 47 consecutive cases of PNSTs were gathered retrospectively from the institutional electronic medical record system (EMR) (dates 2006-2018), comprised of 36 benign and 11 malignant tumors, in total. Inclusion criteria were as follows: adult age (18-75 years), histologic confirmation of lesion, and lesions with MRI protocol containing sequences chosen to be analyzed (i.e. fT2W and T1 pre- and post-contrast images). Exclusion criteria were as follows: incomplete imaging, presence of artifacts limiting evaluation of the lesions, central nervous system location, and pediatric age group. Patient and tumor characteristics are described in Table 1.

Imaging protocol parameters

Imaging protocol parameters were as follows: A) T1W images: Repetition time (TR) / Echo Time (TE) of 600-715 / 9-17 ms, 4 mm slice thickness, with axial and sagittal planes. B) fT2W images: TR/TE of 3600-6000 / 60-62 ms, 4 mm slice thickness, in axial, coronal, and sagittal planes. C) Unenhanced and gadolinium enhanced-3-dimensional (3D) fT1W modified Dixon: TR/TE of 4.6-6.3 / 1.4-1.5, 1.5 mm voxel thickness, in the coronal plane with isotropic
Convolutional Neural Networks Accurately Predict Benign versus Malignant Status Among Peripheral Nerve Sheath Tumors

Mazal et al.

axial and sagittal reconstructions. Contrast administered was intravenous gadolinium at the rate of 0.1 mmol/kg.

**Image pre-processing**

In each tumor, 3D volumes of interest (VOIs) corresponding to tumor boundaries (Figure 2) were manually segmented and contoured by a separate musculoskeletal radiologist (fellowship trained with 4 years of experience) using Velocity software (Varian Medical Systems).

---

**Model development, validation, and testing**

For system training and validation of CNN 1, data extracted from VOIs of f5T2W images of 25 tumors (17 benign, 8 malignant) were used for the proposed CNN, which included 7 convolutional, 3 max-pooling, and 2 fully connected layers (Figure 3). Data augmentation by rotating 3D images and synthetic minority oversampling technique (SMOTE) were employed to balance and increase training samples. CNN model 2 consists of three separate CNN models (same architecture with that of CNN 1) with each trained and validated by VOIs from f5T2W, T1 pre- or post-contrast images of 17 tumors (14 benign, 3 malignant). Data augmentation and SMOTE techniques were employed as well. Each CNN model would generate a predictive malignancy probability for each tumor. Evidential reasoning

---

### Table 1: Characteristics of patients and tumors.

|                      | All Tumors | BPNSTs | MPNSTs |
|----------------------|------------|--------|--------|
| No. of patients      | 47         | 36     | 11     |
| Age (years)          |            |        |        |
| Mean: 41.5           |            | 41.25  | 42.4   |
| Range: 19-75         |            | 19-73  | 21-75  |
| Gender (male/ female)| Male: 17   | Male: 11 | Male: 6 |
|                      | Female: 30 | Female: 25 | Female: 5 |
| Tumor characteristics| Enhancement pattern: Heterogeneous: 25 Homogeneous: 17 (5 tumors without contrast studies) | Enhancement pattern: Heterogeneous: 16 Homogeneous: 16 (4 tumors without contrast studies) | Enhancement pattern: Heterogeneous: 9 Homogeneous: 1 (1 tumor without contrast study) |
|                      | Percent neurofibroma heterogeneously enhancing: 41% (9/22 tumors with contrast studies) | Percent schwannoma heterogeneously enhancing: 70% (7/10 tumors with contrast studies) | Percent MPNST heterogeneously enhancing: 90% (9/10 tumors with contrast studies) |
| Size (max dimension) | Mean: 6.2 cm Range: 0.6 - 25 cm | Mean: 5.6 cm Range: 0.6 - 25 cm | Mean: 8.1 cm Range: 3.1 - 15.9 cm |
| Location             |            |        |        |
| Limb                 | 19         | 17     | 2      |
| Torso                | 15         | 9      | 6      |
| Head & Neck          | 13         | 10     | 3      |
| Histology            |            |        |        |
| Neurofibroma         | 23         | 23     |        |
| Schwannoma           | 13         | 13     |        |
| MPNST                | 11         |        | 11     |
Convolutional Neural Networks Accurately Predict Benign versus Malignant Status Among Peripheral Nerve Sheath Tumors

Mazal et al.

The CNN 2 model predicted benign versus malignant status among PNSTs with an accuracy and AUC of 93%, 0.94, respectively. Blinded radiologists with access to the same imaging dataset (including all imaging data from fsT2W, pre- and post-contrast T1W images mirroring the routine setting) achieved the following accuracies and AUCs: 71%, 0.81 for Radiologist 1, and 71%, 0.70 for Radiologist 2, respectively. No significant differences were found between the accuracies of CNN 2 and Radiologist 1 (p>0.05), CNN 2 and Radiologist 2 (p>0.05), or between Radiologists 1 and 2 (p>0.05). No significant differences were found between the AUCs of CNN 2 and Radiologist 1 (p>0.05), CNN 2 and Radiologist 2 (p>0.05), or Radiologists 1 and 2 (p>0.05).
When comparing the AUCs of Radiologist 1 with access only to fsT2W images versus all imaging data, no significant difference between AUCs was observed (p>0.05). Likewise, in a comparison between the AUCs of Radiologist 2 with access only to fsT2W images versus all imaging data, no significant difference between AUCs was observed, however, this value fell, just short of significance (p=0.0528).

### Discussion

In the comparison of diagnostic performance between the two CNN models and blinded expert radiologists, the CNN models performed comparably or exceeded reader performance in the differentiation of BPNSTs and MPNSTs. The p-values did not reach significance, which may be related to the small sample size. Using fsT2W images in isolation, the accuracies of both CNN 1 and CNN 2 models did not vary significantly from those of either Radiologist 1 or 2 (p>0.05). Interestingly, there was an apparent fall in performance among both radiologists with access to all imaging data, and their performance appeared to be better using only fsT2W images. This could possibly be attributed to the heterogeneous enhancement patterns observed in some benign tumors. Such tumor characteristics are well-described and are not uncommon in the clinical setting, leading to diagnostic dilemmas [4, 17-19]. Heterogeneous enhancement and intra-tumoral necrosis are particularly common findings among ancient schwannomas; however, they can also be seen sometimes in the setting of other BPNSTs such as plexiform neurofibromas [19]. Within this study, 70% of benign schwannomas and 41% of neurofibromas exhibited heterogeneous enhancement. Larger dimensions (ranging up to 25 cm) were also commonly observed among benign lesions, particularly among patients with neurofibromatosis Type 1 (NF-1). Likewise, these features are commonly observed with MPNSTs. MPNSTs have been reported to differ from BPNSTs by their larger diameter (> 5 cm), peri-lesional edema, peripheral or heterogeneous enhancement, and intra-tumoral necrosis or hemorrhage [20]. Although in combination, these features confer a high specificity for malignancy, such features are insensitive and are not present in all MPNSTs [20]. Given the presence of many of these features in benign tumors, CNNs may serve as a particularly useful diagnostic tool with PNSTs demonstrating borderline or intermediately concerning imaging characteristics. It may also help reduce the number of unnecessary biopsies among patients with NF-1, in whom larger benign lesions with inhomogeneous enhancement are particularly common.

The study was limited by a small sample size, as only histology-proven diagnoses were included; however, study subjects were identified randomly and are believed to represent a sufficiently representative sample to render these results generalizable and reproducible for other medical centers. Future studies with larger testing datasets will be essential to increase the power sufficiently to discern more subtle differences in performance between CNN models and expert readers. Furthermore, due to the presence of imbalanced classes within the testing datasets, which heavily favored representation of benign lesions, there were additional limitations on the statistical inferences which could be drawn herein. The presence of a minority class (MPNSTs) in the dataset predisposes to misclassification and is inappropriate in the setting of machine learning algorithms set to evaluate accuracy, as it may yield poor predictive accuracy in the minority class. To reconcile this imbalance, SMOTE was utilized to oversample the minority class and balance the dataset, with good resulting accuracy and AUC, as observed by both CNN models.

Because CNN models were utilized in this study, we did not seek to extract specific radiomic features. Future studies with larger cohorts may help further delineate individual clinical biomarkers indicative of malignancy. Further studies will also be necessary to determine whether CNNs or radiomics can play a role in the differentiation of BPNST subtypes, which was not pursued herein.

Finally, an additional limitation of this study was that radiologists were not given access to clinical history or other supplementary data that would ordinarily be found in the routine setting, as we sought to specifically compare unbiased reader performance to the machine reads.

### Conclusion

In conclusion, the results of this study demonstrate that CNNs differentiate benign and malignant PNSTs with high accuracy, particularly using multiple imaging sequences in combination modeling the routine setting. In future clinical practice, CNNs are most likely to serve as a useful tool in differentiating PNSTs with borderline imaging features concerning for malignancy.

### References

1. Levi AD, Ross AL, Cuartas E, Qadir R, Temple HT. 2010. The surgical management of symptomatic peripheral nerve sheath tumors. *Neurosurgery* 66(4): 833-840. [https://doi.org/10.1227/01.NEU.0000367636.91555.70](https://doi.org/10.1227/01.NEU.0000367636.91555.70)

2. Farid M, Demicco EG, Garcia R, Ahn L, Merola PR, et al. 2014. Malignant peripheral nerve sheath tumors. *Oncologist* 19(2): 193-201. [https://doi.org/10.1634/theoncologist.2013-0328](https://doi.org/10.1634/theoncologist.2013-0328)
Convolutional Neural Networks Accurately Predict Benign versus Malignant Status Among Peripheral Nerve Sheath Tumors

Mazal et al.

3. Kim DH, Murovic JA, Tiel RL, Moes G, Kline DG. 2005. A series of 397 peripheral nerve sheath tumors: 30-year experience at Louisiana State University Health Science Center. J Neurosurg 102(2): 246-255. https://doi.org/10.3171/jns.2005.102.2.0246

4. Demehri S, Belzberg A, Blakeley J, Fayad LM. 2014. Conventional and functional MR imaging of peripheral nerve sheath tumors: initial experience. AJNR Am J Neuroradiol 35(8): 1615-1620. https://doi.org/10.3174/ajnr.A3910

5. Shahid KR, Amranii KK, Esther RJ, Lovey JV, Spinner RJ. 2011. False-negative fluorine-18-fluorodeoxyglucose positron emission tomography of a malignant peripheral nerve sheath tumor arising from a plexiform neurofibroma in the setting of neurofibromatosis type 1. J Surg Orthop Adv 20(2): 132-135

6. Zhou M, Scott J, Chaudhury B, Hall L, Goldgof D, et al. 2018. Radiomics in brain tumor: image assessment, quantitative feature descriptors, and machine-learning approaches. AJNR Am J Neuroradiol 39(2): 208-216. https://doi.org/10.3174/ajnr.A5391

7. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, et al. 2017. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14(12): 742-762. https://doi.org/10.1038/nrclinonc.2017.141

8. Aerts H, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, et al. 2014. Decoding tumor phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5: 4006. https://doi.org/10.1038/ncomms5006

9. van Rossum PS, Fried DV, Zhang L, Hofstetter WL, van Vulpen M, et al. 2016. The incremental value of subjective and quantitative assessment of 18F-FDG PET for the prediction of pathologic complete response to preoperative chemoradiotherapy in esophageal cancer. J Nucl Med 57(5): 691-700. https://doi.org/10.2967/jnumed.115.163766

10. Coroller TP, Grossman P, Hou Y, Velazquez ER, Leijenaar RTH et al. 2015. CT-based radiomic signatures predict distant metastasis in lung adenocarcinoma. Radither Oncol 114(3): 345-50. https://doi.org/10.1016/j.radonc.2015.02.015

11. Liang C, Huang Y, He L, Chen X, Ma Zelan, et al. 2016. The development and validation of a CT-based radiomics signature for the preoperative discrimination of stage I-II and stage II-IV colorectal cancer. Oncotarget 7(21): 31401-3112. https://doi.org/10.18632/oncotarget.8919

12. Hawkins S, Wang H, Liu Y, Garcia A, Stringfield O, et al. 2016. Predicting Malignant Nodules from Screening CT Scans. J Thorac Oncol 11(12): 2120-2128. https://doi.org/10.1016/j.jtho.2016.07.002

13. Grossman P, Gutman DA, Dunn WD Jr, Holder CA, Aerts HJWL, et al. 2016. Imaging-genomics reveals driving pathways of MRI derived volumetric tumor phenotype features in Glioblastoma. BMC Cancer 16: 611. https://doi.org/10.1186/s12885-016-2659-5

14. Huang Y, Liu Z, He L, et al. 2016. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. Radiology 281(3): 947-957. https://doi.org/10.1148/radiol.2016152234

15. Li H, Zhu Y, Burnside ES, Drukker D, Hoadley KA, et al. 2016. MR imaging radiomics signatures for predicting the risk of breast cancer recurrence as given by research versions of mammaprint, oncotypex DX, and PAM50 gene assays. Radiology 281(2): 382-391. https://doi.org/10.1148/radiol.2016152110

16. DeLong ER, DeLong DM, Clarke-Pearson DL. 1988. Comparing the areas under two or more related correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44(3): 837-45

17. Hanemann CO, Evans DG. 2006. News on the genetics, epidemiology, medical care and translational research of Schwannomas. J Neurol 253(12): 1533-1541. https://doi.org/10.1007/s00415-006-0347-0

18. Mazal AT, Ashikyan O, Cheng J, et al. 2018. Diffusion-weighted imaging and diffusion tensor imaging as adjuncts to conventional MRI for the diagnosis and management of peripheral nerve sheath tumors: current perspectives and future directions. Eur Radiol 28(9): 4123-4132. https://doi.org/10.1007/s00330-018-5838-8

19. Ahlawat S, Blakeley J, Montgomery E, Subramaniam RM, Belzberg A, et al. 2013. Schwannoma in neurofibromatosis type 1: a pitfall for detecting malignancy by metabolic imaging. Skeletal Radiol 42(9): 1317-1322. https://doi.org/10.1007/s00256-013-1626-3

20. Wasa J, Nishida, Tsukushi S, Shido Y, Sugura H, et al. 2010. MRI features in the differentiation of malignant peripheral nerve sheath tumors and neurofibromas. AJR Am J Roentgenol 194(6): 1568-1574. https://doi.org/10.2214/AJR.09.2724