Neuro-Oncology
24(10), 1790–1798, 2022 | https://doi.org/10.1093/neuonc/noac100 | Advance Access date 15 April 2022

Evaluation of magnetic resonance imaging-based radiomics characteristics for differentiation of benign and malignant peripheral nerve sheath tumors in neurofibromatosis type 1

Inka Ristow, Frederic Madesta, Lennart Well, Farzad Shenas, Felicia Wright, Isabel Molwitz, Said Farschtschi, Peter Bannas, Gerhard Adam, Victor F. Mautner, René Werner†, and Johannes Salamon†

Department of Diagnostic and Interventional Radiology and Nuclear Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany (I.R., L.W., F.S., F.W., I.M., P.B., G.A., J.S.); Institute of Computational Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany (F.M., R.W.); Department of Neurology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany (S.F, V.F.M.)

†These authors contributed equally to this work.

Corresponding Author: Inka Ristow, MD, Department of Diagnostic and Interventional Radiology and Nuclear Medicine, University Medical Center Hamburg-Eppendorf, Martinistraße 52, Hamburg 20246, Germany (i.ristow@uke.de).

Abstract

Background. Patients with neurofibromatosis type 1 (NF1) develop benign (BPNST), premalignant atypical (ANF), and malignant (MPNST) peripheral nerve sheath tumors. Radiological differentiation of these entities is challenging. Therefore, we aimed to evaluate the value of a magnetic resonance imaging (MRI)-based radiomics machine-learning (ML) classifier for differentiation of these three entities of internal peripheral nerve sheath tumors in NF1 patients.

Methods. MRI was performed at 3T in 36 NF1 patients (20 male; age: 31 ± 11 years). Segmentation of 117 BPNSTs, 17 MPNSTs, and 8 ANFs was manually performed using T2w spectral attenuated inversion recovery sequences. One hundred seven features per lesion were extracted using PyRadiomics and applied for BPNST versus MPNST differentiation. A 5-feature radiomics signature was defined based on the most important features and tested for signature-based BPNST versus MPNST classification (random forest [RF] classification, leave-one-patient-out evaluation). In a second step, signature feature expressions for BPNSTs, ANFs, and MPNSTs were evaluated for radiomics-based classification for these three entities.

Results. The mean area under the receiver operator characteristic curve (AUC) for the radiomics-based BPNST versus MPNST differentiation was 0.94, corresponding to correct classification of on average 16/17 MPNSTs and 114/117 BPNSTs (sensitivity: 94%, specificity: 97%). Exploratory analysis with the eight ANFs revealed intermediate radiomic feature characteristics in-between BPNST and MPNST tumor feature expression.

Conclusion. In this proof-of-principle study, ML using MRI-based radiomics characteristics allows sensitive and specific differentiation of BPNSTs and MPNSTs in NF1 patients. Feature expression of premalignant atypical tumors was distributed in-between benign and malignant tumor feature expressions, which illustrates biological plausibility of the considered radiomics characteristics.

Key Points

- Radiomics allows for sensitive and specific differentiation between BPNSTs and MPNSTs.
- ANFs show intermediate radiomic feature characteristics between BPNSTs and MPNSTs.

Importance of the Study

Malignant transformation of peripheral nerve sheath tumors represents a major life-limiting factor for patients with neurofibromatosis type 1 (NF1). Therefore, early detection of malignant peripheral nerve sheath tumors and their premalignant precursor, atypical neurofibromas, is of great importance. The exclusive use of morphologic tumor characteristics determined by magnetic resonance imaging (MRI) has limited diagnostic accuracy. We, therefore, assessed the value of radiomics for automatic classification of benign, atypical, and malignant peripheral nerve sheath tumors in NF1 patients. In summary, MRI-based radiomics allows sensitive and specific distinction of benign versus malignant, and of benign versus (pre-)malignant peripheral nerve sheath tumors in NF1 patients. Hence, MRI-based radiomics machine learning classification represents a promising noninvasive diagnostic radiological imaging tool to be investigated in a larger study population in the near future.

Neurofibromatosis type 1 (NF1) is an autosomal-dominantly inherited neurogenetic disorder caused by a mutation of the cell growth regulating protein neurofibromin in the NF1 gene at 17q11.2. With an incidence of about 1:2500–1:3000 NF1 is considered one of the most common hereditary neurocutaneous diseases.[1,2]

Typically, patients have a predisposition to develop benign peripheral nerve sheath tumors (BPNSTs) of the skin (= cutaneous neurofibromas), the subcutis, and the deeper soft tissue (= plexiform neurofibromas). The latter are composed of Schwann cells, fibroblasts, mast cells, perineurial cells, and a rich network of collagen fibers.[3] Based on their morphological appearance, different subtypes of internal BPNSTs have been described, namely discrete (involving only one nerve fascicle) and plexiform (involving multiple nerve fascicles) neurofibromas. Due to their increased risk of malignant transformation, plexiform neurofibromas are of particular clinical relevance.[4] Malignant tumors develop in about 8%–16 % of NF1 patients during lifetime.[5,6] Since malignant peripheral nerve sheath tumors (MPNSTs) tend to metastasize early and are often resistant to chemotherapy, their occurrence serves as the most important life-limiting factor in NF1 patients.[7] Therefore, early detection and resection is of high relevance.

Atypical Neurofibromas (ANFs) were first described in 2011 by Beert et al.[9] as a histopathological intermediate tumor type typically located in the deep soft tissue. As premalignant precursors to MPNSTs, ANFs show neither local recurrence after previous resection nor the ability to metastasize.[8,9] Therefore, early, sensitive, and specific noninvasive diagnosis of (pre-)malignant nerve sheath tumors is of vital importance for clinical management of NF1 patients.

Magnetic resonance imaging (MRI) serves as the imaging modality of choice for long-term monitoring of affected NF1 patients with a fat-saturated T2-weighted sequence considered as the pivotal MRI sequence.[10] In recent years, there has been an increased research interest on quantitative imaging parameters for better risk stratification of neurofibromas. For example, a correlation was found between high total tumor burden and risk for malignant transformation.[11–13] MRI-based "worrisome features" include lobulated appearance, irregular tumor margin contours, peritumoral edema, and intratumoral heterogeneity including intratumoral hemorrhages or cystic changes.[14–16] In addition to morphological MR imaging, the use of diffusion weighted imaging (DWI)[17] and alternative imaging modalities, such as positron emission tomography (PET)[18–20] adds to a more precise tumor characterization by providing information about local tumor architecture and tumoral metabolism.

Radiomics allows detailed characterization and analysis of structures in radiological images by using quantitative image features (eg, textures or homo-/heterogeneity of signal intensities), which are frequently not visible to even radiologically trained eyes. In combination with machine-learning (ML) techniques, radiomics allows automatic extraction and differentiation of typical radiomics signatures or feature expressions.[21] Uthoff et al.[22] investigated the potential of radiomic features derived from MRI and PET/CT data for the differentiation of BPNSTs and MPNSTs. However, they did not use an ML algorithm and did not include patients with premalignant ANFs.

The purpose of this study was to evaluate the value of an MRI-based radiomics ML classifier for automatic classification of BPNST and MPNST in NF1 patients and to explore the plausibility of a respective radiomics signature by further consideration of premalignant ANFs.

Materials and Methods

This single-center retrospective data evaluation was approved by the local ethics board with waiver of informed consent (WF-039/21). All procedures complied with the local data protection guidelines as well as the Declaration of Helsinki.

Study Population

Inclusion criteria were diagnosis of NF1 according to the NIH criteria[23] and availability of high resolution 3T MRI examinations with axial T2-SPAIR weighted MRI sequences. MRI studies were acquired between August 2014 and April 2021.

The study population included 36 adolescents and adults (20 male; mean age 31.3 years; range 12–54 years) (Table 1). The total data set included 117 BPNSTs, 17 MPNSTs, and 8 ANFs. Tumors were considered benign when no changes in size or appearance were present in follow-up examinations within ≥ 24 months. All malignant tumors were characterized by size progression. Regarding the ANFs, seven out of eight patients presented to our institution due to tumor size progression and indication for further diagnostics to

Table 1. Demographic and Clinical Characteristics of the NF1 Study Population

Parameter	Value
Subjects (*n*)	36
Age (years, mean ± SD)	31.3 ± 11.3
(range)	12–54
Sex (m/f)	20/16
Tumor type (*n*)	
BPNST	117
Plexiform/discrete	75/42
ANF	8
MPNST	17
Tumor volume (mm^3, mean ± SD)	
BPNST	29017 ± 70894
ANF	60554 ± 40990
MPNST	269159 ± 268774

SD, standard deviation; m, male; f, female; BPNST, benign peripheral nerve sheath tumor; MPNST, malignant peripheral nerve sheath tumor; ANF, atypical neurofibroma.

exclude malignant transformation. For the eighth patient no previous imaging was available, because NF1 was diagnosed for the first time.

Resected MPNSTs were classified according to the grading system of the Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC).[24] Histopathological evaluation after tumor resection served as the reference standard for MPNSTs and ANFs. Histopathological reports following MPNST resection revealed 1 FNCLCC grade I, 6 FNCLCC grade II, and 10 FNCLCC grade III tumors. Considering the topographic distribution of the tumors, we observed 29 cervical tumors (20.4%; 24 BPNSTs; 3 MPNSTs; 2 ANFs), 20 tumors of the thorax/axilla (14.1%; 11 BPNSTs; 7 MPNSTs; 2 ANFs), 29 tumors of the abdomen/retroperitoneum (20.4%; 27 BPNSTs; 0 MPNSTs; 2 ANFs), 54 tumors of the pelvis (38.0%; 51 BPNSTs; 1 MPNST; 2 ANFs), 9 tumors of the lower extremities (6.3%; 4 BPNSTs; 5 MPNSTs; 0 ANFs), and 1 MPNST located at the upper extremity (0.7%).

MRI Data Acquisition

MR imaging was performed at 3T (Philips Ingenia, Best, The Netherlands). The imaging protocol included a localizer, a coronal T2-weighted turbo spin-echo sequence, and an axial T2-weighted turbo spin-echo sequence with fat suppression (spectral attenuated inversion recovery, SPAIR). Subsequent analyses were based on the SPAIR sequences (TR 3822 ms, TE 80 ms, flip angle 90°, matrix 244 × 215, FOV 270 × 270 mm, voxel size 0.7 × 0.7 mm, slice thickness 3 mm, intersection gap 0 mm).

Tumor Segmentation

MRI data were converted from DICOM into NIfTI using SPM12 (Wellcome Trust Centre for Neuroimaging, London, United Kingdom) running on Matlab (MathWorks, Inc., Natick, Massachusetts, United States; Version R2020a). Tumor segmentation was performed manually on the T2-weighted SPAIR images using ITK-SNAP 3.8.0[25] by a radiologist (I.R.) with 3 years of expertise in MRI. Segmented data sets were reviewed and, if necessary, manually edited by a senior physician (J.S.) with 10 years of expertise in MRI and a research focus on multimodal imaging techniques in neurofibromatosis spectrum disorders. Exemplary MRI-based segmentations of a BPNST, ANF, and MPNST are displayed in Figure 1.

ML Analysis Strategy

Radiomics-based differentiation of BPNST and MPNST was subdivided into two parts. In the first part, the respective two-class classification problem BPNST versus MPNST was considered and general feasibility was determined by definition of a characteristic signature. In the second part, based on the signature, we analyzed signature feature expressions of ANF in comparison to BPNST and MPNST and explored feasibility of ML-based differentiation of BPNST, ANF, and MPNST.

Radiomics-based differentiation of MPNSTs and BPNSTs

A standard set of quantitative image features according to the PyRadiomics Python package, version 3.0.1[26] was computed for the individual lesion segmentation masks and corresponding T2-weighted SPAIR images. The feature set comprised 18 first-order intensity, 14 shape, and 75 texture features (107 features in total); due to the limited sample size of this proof-of-principle study, wavelet decompositions were not performed to reduce the initial feature set size.

In a first step, random forests (RFs; Python scikit-learn, version 0.24.2)[27] were trained to differentiate BPNST and MPNST samples based on the entire 107 features. RFs are ML algorithms with comparably low sensitivity to class imbalance, able to handle large numbers of heterogeneous predictors and cluster-correlated observations (here: patients with multiple lesions).[28] This first step primarily aimed at identification of relevant features. RF outputs were evaluated by leave-one-patient-out cross-validation (LOOCV). To account for RF randomness, LOOCV was repeated 10 times. Feature importance was evaluated by averaging the feature-specific Gini importance[26] for the 134 × 10 = 1340 RFs of the repeated LOOCV. To double-check plausibility of feature importance values, a random variable (generated using a normal random number generator) was added to the initial feature set and its "importance" compared to corresponding image feature values.[29]

In a second step, a problem-specific radiomics signature for differentiation of BPNSTs and MPNSTs was defined that consists of a small number of relevant image features. The signature feature number was determined by analysis of the mean corresponding area under the receiver operator characteristic (ROC) curve (AUC) obtained for RF-based classification with different feature sets (starting with only the single most important feature, and successively adding further features according to their importance values derived in the first part of the experiments; evaluation again

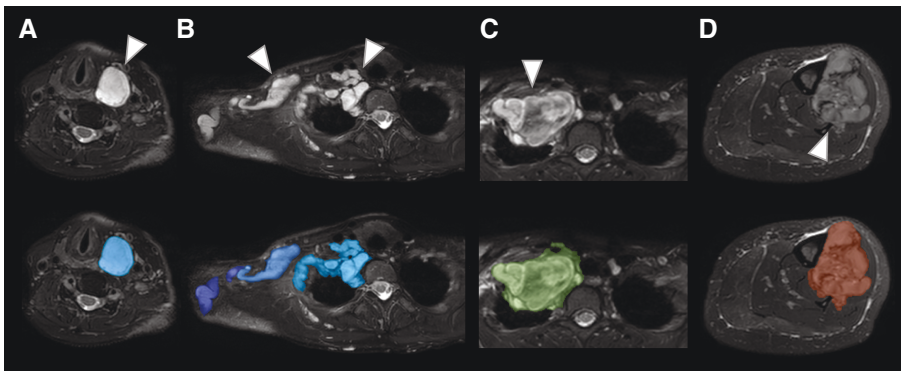

Figure 1. Exemplary MRI-based segmentations of benign BPNST, premalignant ANF, and malignant MPNST. (A) Benign discrete cervical BPNST with a nodular appearance (arrowhead) in a 39-year-old woman. (B) Benign plexiform BPNST of the neck and shoulder (arrowheads) in a 37-year-old woman. (C) Premalignant ANF of the neck and shoulder (arrowhead) in an 18-year-old woman. (D) MPNST of the tibia (arrowhead) in a 41-year-old woman. Top row: transverse slices of a T2 SPAIR sequence at 3T; lower row: overlay with indicated segmentation of tumors.

by means of repeated LOOCV). The smallest set of most important features that achieved AUC values in saturation (no additional value of adding further features) defined the problem-specific radiomics signature, for which the classification performance was evaluated in detail.

All experiments were based on RFs with 100 trees. Confusion matrices, as well as sensitivity and specificity values, correspond to an optimal operating point determined according to Youden's index. For improved interpretability, the median image feature values were removed, and the data rescaled according to the feature-specific interquartile range.

ANF feature expression and differentiation of BPNST, ANF, and MPNST

We hypothesize that extracted radiomics signature is at least partly characteristic for the underlying biology of peripheral nerve sheath tumors. Therefore, feature expressions of ANFs as premalignant precursors to MPNSTs should be in between BPNST and MPNST feature expressions. This hypothesis was tested as a first step. Moreover, the potential of the radiomics signature for RF-based differentiation between BPNSTs, ANFs, and MPNSTs was investigated. Considering the small sample sizes of the ANFs and MPNSTs, we resorted to the following two-class problems: 1) ANF versus BPNST, 2) ANF or MPNST versus BPNST (ie, identification of nerve sheath tumors in need of further clarification), and 3) MPNST versus ANFs. RF training was based on only the image features of the defined signature. The validation strategy (repeated LOOCV) and algorithm configuration were similar to the BPNST versus MPNST experiments.

Results

Radiomics-Based Differentiation of BPNST and MPNST

Highest performance for differentiation of BPNST and MPNST was achieved with a five features RF classification with a ROC AUC value of 0.94 (mean of the repeated LOOCV runs; standard deviation (SD) < 0.01). Classification based on more than the five most important features did not lead to further performance gain, and the combination of these features was considered the sought radiomics signature for differentiation of BPNSTs and MPNSTs. The AUC values for RF-based differentiation of BPNST and MPNST adopting a different number of quantitative image features (features successively added according to their RF feature importance) as well as feature importance values can be found in Figure 2. The figure illustrates that for five or more features RF classification performance is stable. All five signature features were texture features, computed based on the gray level co-occurrence matrix (GLCM), namely: inverse difference moment normalized, correlation, inverse difference normalized, maximal correlation coefficient, and difference average (see van Griethuysen et al.[26] for details). RF classification based on the 5-feature signature led to, on average, correct classification of 16 of 17 MPNSTs and 114 of 117 BPNSTs, corresponding to a sensitivity of 94% and a specificity of 97% (mean over LOOCV runs; both SD < 1%). The RF predictions were very stable for the repeated runs. In 2/10 runs 113/117 BPNSTs were correctly classified; in the remaining runs, 114/117. The numbers of the correctly assigned MPNSTs did not vary. The ROC curves and the mean confusion matrix for the repeated LOOCV runs are shown in Figure 3.

ANF Feature Expression and Differentiation of BPNST, ANF, and MPNST

Analysis of the ANF feature expression showed a consistent distribution in between the BPNST and MPNST features for all five most relevant features for the differentiation of BPNST and MPNST. Thus, image feature expressions of the ANFs correspond to the ANF description as intermediate tumor type and premalignant precursor of MPNST. The distributions of the feature expressions for the five features of the defined radiomics signature and the BPNST, ANF, and MPNST groups are shown in Figure 4. Distributions are

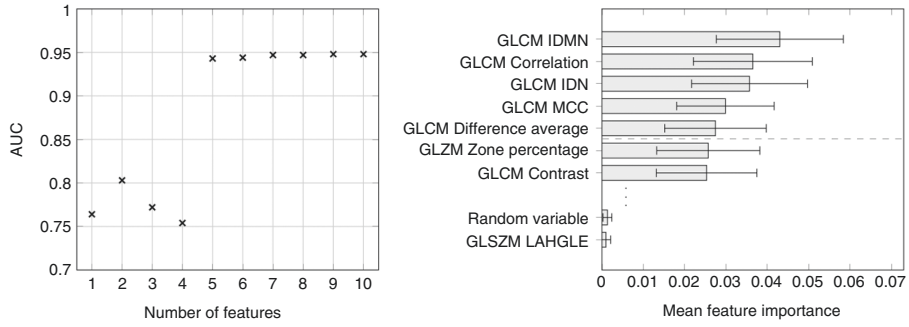

Figure 2. Area under the receiver operator characteristic curve (AUC) for random forest-based differentiation and radiomic feature importance. (Left) Area under the receiver operator characteristic curve (AUC) for random forest-based differentiation of MPNSTs and BPNSTs when adopting the 1, 2, … up to 10 most important radiomics features of the repeated leave-one-patient-out cross-validation (LOOCV) of the classification based on the entire initial feature set (107 radiomics features). Combining the five most important figures for classification stabilizes the AUC (ROC AUC value about 0.95); adding more features does not lead to further gain of performance. (Right) Importance of the radiomics features in the initial RF experiments. The five most important features, which were subsequently considered the radiomics signature, were all computed based on the gray level co-occurrence matrix (GLCM). The random variable was added as a plausibility check; the respective "importance" is, as expected, lower than that of most (ie, 106/107) of the considered radiomics features. GLCM, gray level co-occurrence matrix; IDMN, inverse difference moment normalized; IDN, inverse difference normalized; MCC, maximal correlation coefficient; GLSZM, gray level size zone matrix; LAHGLE, large area gray level emphasis.

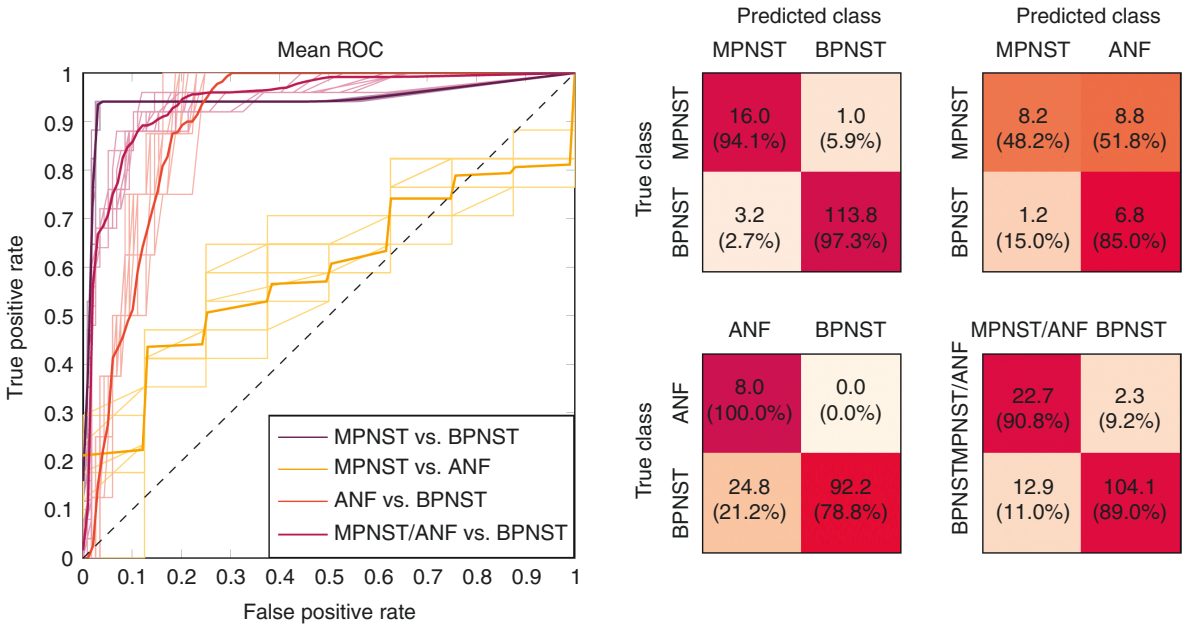

Figure 3. ROC curves and confusion matrices for the considered classification problems. (Left) ROC curves for the considered classification problems: MPNST versus BPNST, ANF versus BPNST, MPNST versus ANF, MPNST/ANF versus BPNST. The bold lines represent the mean ROC curves of the repeated LOOCV runs, and the thin lines the corresponding individual runs. (Right) The confusion matrices show the mean numbers (absolute and, in parentheses, corresponding percentage values) of the repeated runs, evaluated for an optimal operating point according to Youden's index.

estimated by kernel density estimation based on the observations of the present data cohort according to Scott.[30]

The partially overlapping feature expression distributions of the three classes led, however, to decreased classification performance when considering ANFs. ROC curves and mean confusion matrices for the respective classification problems are shown in Figure 3. Automated differentiation of ANFs and BPNSTs resulted in an AUC of

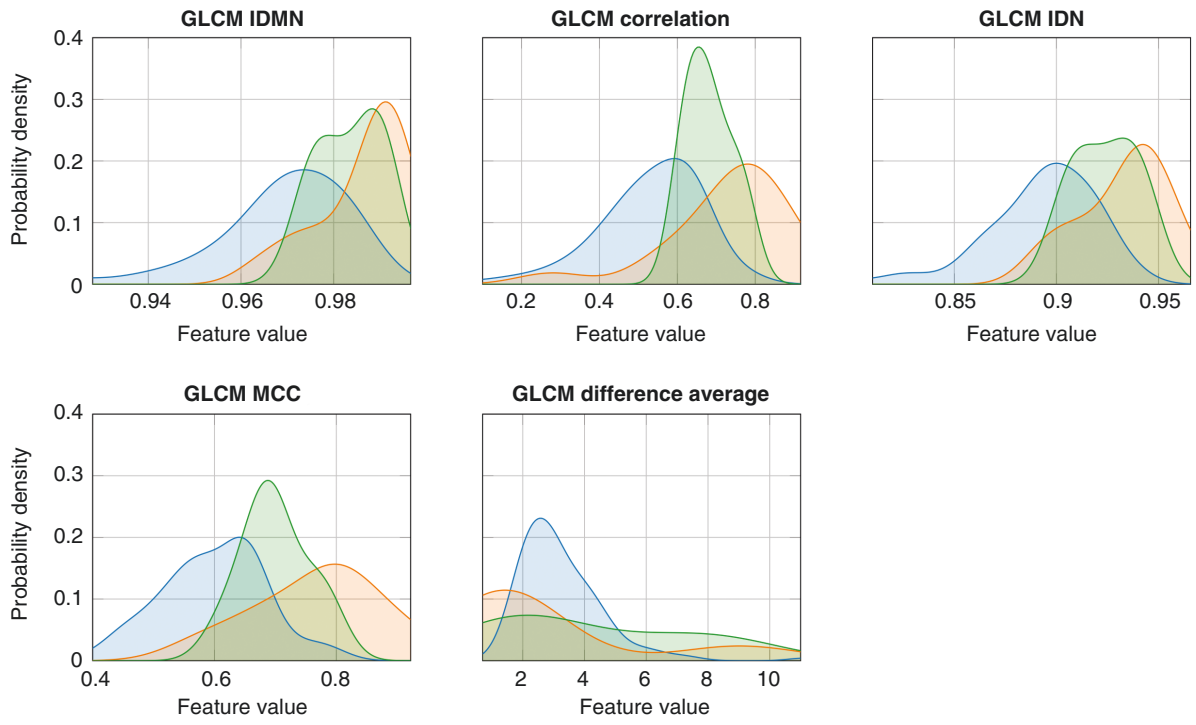

Figure 4. Distributions of the defined radiomics signature features of the BPNSTs, the MPNSTs, and ANFs. In line with the clinical interpretation of ANFs as a histopathological premalignant precursor, the distributions of the ANF feature (green) lie in-between those of BPNSTs (blue) and MPNSTs (red). GLCM, gray level co-occurrence matrix; IDMN, inverse difference moment normalized; IDN, inverse difference normalized; MCC, maximal correlation coefficient.

0.91 ± 0.01, corresponding to correct classification of all ANFs for all LOOCV runs and, on average, 92/117 BPNSTs (sensitivity: 100% ± 0%; specificity: 79% ± 3%); differentiation of the small sample size classes MPNST and ANF yielded correct classification of 7/8 ANFs, but only 8/17 MPNSTs for the selected operating point (AUC: 0.60 ± 0.02; sensitivity: 48% ± 13%; specificity: 85% ± 9%). For the differentiation of premalignant or malignant tumors and BPNSTs, on average, 23 of the 25 lesions "in need of further clarification" (ie, MPNSTs or ANFs) were correctly classified as such, and 104/117 of the benign tumors were classified as BPNSTs (AUC: 0.94 ± 0.01; sensitivity: 91% ± 3%; specificity: 89% ± 2%).

Discussion

In this proof-of-principle study, we demonstrated the feasibility of an MRI-based radiomics ML classifier for automatic differentiation of BPNST from MPNST in patients with NF1. However, differentiation of MPNSTs and premalignant ANFs was not reliably possible due to overlapping feature expression distributions. RF-based classification adopting a 5-feature radiomics signature resulted in automated MPNST and BPNST differentiation with a 94% sensitivity and 97% specificity. Including ANFs and combining MPNSTs and ANFs to a single class resulted in an automated differentiation from BPNSTs with a 91% sensitivity and 89% specificity, indicating feasibility of radiomics-based identification of neurofibromatosis tumors in need for further clarification, for example, by performing a PET/CT or biopsy.

This single-center study indicates that it is feasible to automatically classify BPNST and MPNST based on quantitative radiomic features derived from non-enhanced MRI and that the derived diagnostic accuracy is comparable to that of existing NF1 studies. As recently reported in a meta-analysis by Martin et al.[31], heterogeneous pooled diagnostic accuracies were observed with respect to MRI-based morphological imaging features for the differentiation of BPNSTs versus MPNSTs (33%–99% sensitivity, 33%–95% specificity) with absence of target sign yielding highest sensitivity. DWI-based differentiation of BPNSTs and MPNSTs revealed comparably high diagnostic accuracies with 92% sensitivity and 98% specificity.[17] Regarding PET/CT pooled sensitivities and specificities yielded similar accuracies of standardized uptake value (SUV_{max}) (94% sensitivity, 81% specificity) and tumor-to-liver ratio (93% sensitivity, 79% specificity), whereas best accuracy was achieved for $SUV_{max} \geq 3.5$ median cutoff range (99% sensitivity, 75% specificity).[31] However, one disadvantage of PET/CT is the associated radiation exposure, especially in this particularly vulnerable collective of young adults. Our MRI-based radiomics BPNST versus MPNST classification with 94% sensitivity

and 97% specificity is therefore comparable to other approaches.

The difficulties we often face when differentiating MPNSTs from ANFs on MRI are due to the nature of these tumors, for example, because of focal malignancy in a large PNF or overlapping feature expression. This is similarly reflected in the challenging histopathological differentiation.[8] Despite morphological and immunohistochemical diagnostic criteria, the distinction of ANF from (low-grade) MPNST may pose some diagnostic difficulties, making diagnosis for even experienced pathologists challenging. Criteria for the diagnosis ANF are met when at least two worrisome features are present but histologically fall short of MPNST as, for instance, nuclear atypia, loss of the CD34+ fibroblastic network, hypercellularity, and an elevated mitotic index (>1/50 high power fields (HPF) and < 3/10 HPF).[8] Other immunohistochemical markers, such as S100, CD34, Ki67 or of p53 as well as the deletion of CDKN2A/B on chromosome 9 (gene locus p21.3) may allow reliable conclusions to be drawn.[9,32,33] Therefore, the fact that there is also significant overlap in the radiomics feature expression of these two entities, in particular, underscores the plausibility of the distribution of the derived radiomics signatures.

In our study, the feasibility of radiomics for NF1 tumors was performed based on fat-suppressed T2-weighted sequence, which serves as the most important MRI sequence for diagnostics in NF1 patients.[10] Besides, other MRI sequences and imaging modalities have been suggested as diagnostically helpful for NF1 tumor differentiation, for example, DWI or PET/CT.[17,19,20] A texture analysis-based approach by Cook et al.[34] provided evidence that heterogeneity and higher-order features derived from ^{18}F-FDG-PET/CT data can be applied for differentiation of BPNST and MPNST. However, like in the study by Uthoff et al.[22], these authors did not apply ML techniques for automatic tumor classification. Future studies should evaluate feasibility of radiomics for these different techniques and answer the question whether a combination of these techniques towards a multimodal/complementary radiomics approach improves classification accuracy of NF1 tumors. Moreover, it would be of particular interest to correlate the therapeutic course of non-resectable PNFs under MEK inhibitor selumetinib therapy with radiomics-based feature expression for the assessment or prediction of therapy response, since previous studies have shown that selumetinib not only stops tumor growth but even leads to a tumor size reduction.[35]

In this study, we exclusively evaluated texture parameters. In contrast to morphological features such as size and shape, a texture-based approach might allow better generalizability due to less susceptibility to data acquisition-related constraints. For example, during MRI data acquisition large plexiform neurofibromas, which can often be delineated along the entire plexus, are not completely captured within one stack in a cranial or caudal direction. Artifact-free fusion of these stacks is also not always possible.

Analysis of tumor volumes extracted from segmentation revealed that malignant tumors were significantly larger than benign tumors (Wilcoxon rank-sum test, $P < .001$; Table 1). This observation is in line with previous work. For example, Martin et al.[31] reported a pooled accuracy of 0.71 sensitivity and 0.85 specificity in their meta-analysis with respect to the tumor size criterion which was derived from investigations from Ahlawat et al., Karsy et al., and Matsumoto et al.[36–38] To investigate how well tumor size can discriminate between benign and malignant nerve sheath tumors a similar methodological setup was used to train a logistic regression model to differentiate between MPNST and BPNST based on only the tumor size. The corresponding ROC AUC was 0.69 (sensitivity: 65%, specificity: 92%), hence lower compared to the texture-based radiomics signature. Premalignant tumors were also significantly larger than benign tumors (Wilcoxon rank-sum test, $P < .001$; Table 1). The AUC for differentiation between ANF and BPNST based on only the tumor size was 0.75 (sensitivity: 88%, specificity: 66%) and for differentiation of lesions in need for further classification (ANF, BPNST) and benign tumors 0.72 (sensitivity: 64%, specificity: 88%). Again, the values are lower compared to radiomics signature-based differentiation. Size differences between premalignant and malignant tumors were not significant in our population ($P = .26$).

Our study has the following limitations: first, small sample size and a correspondingly small number of target lesions for radiomics is a well-known general concern,[39] especially for diseases with low prevalence such as NF1. Therefore, a "classical" radiomics approach with a separate training and test data set was not feasible, which is why a LOOCV suitable for smaller collectives was applied in this proof-of-principle study. One consequence to overcome this problem in the future may be the establishment of multisite databases, which would allow further subgroup analyses considering other relevant factors such as age and sex. The proposed radiomics signature could also be evaluated for data sets with different technical acquisition parameters (eg, different field strength). A multimodal approach considering other patient-specific and disease-relevant risk factors for malignant transformation, such as genetics, total tumor burden, and other clinical parameters[11–13,40] might help to optimize personalized treatment and risk stratification by means of radiomics and ML. For example, in a radiogenomics study, Liu et al.[41] found significant associations of NF1 mutations and radiomic MRI features in a genotype-phenotype correlation. In accordance with the previously mentioned limitation, the rarity of this disease and even rarer occurrence of malignant transformation of peripheral nerve sheath tumors results in a particular low number of target lesions of malignant (MPNSTs) as well as premalignant (ANFs) lesions. Therefore, having only 17 MPNSTs and 8 ANFs in our data set, we cannot completely rule out overfitting. A second limitation is the class imbalance, although and as already mentioned in the methods section, the applied RF model has a comparably low sensitivity to class imbalance and is also able to handle cluster-correlations.[28] The third limitation is the performed manual tumor segmentation approach which implies a certain degree of observer dependence within the ML process. However, it has been shown that radiomic feature expressions including GLCM features as applied in this study are comparably stable regarding variations in segmentations.[42] We have deliberately decided to not use a semi-automatic segmentation approach, as this did

not seem to work effectively in our hands. For example, threshold-based segmentation did not succeed in a correct differentiation of nerves and adjacent vessels. Similarly, when using semi-automatic segmentation, it is not possible to exclusively segment the peripheral nerve without sparing the adjacent spinal cord. Likewise, threshold-based segmentation was also not promising in the pelvis since the ovaries and seminal vesicles were included in the target regions of interest due to similar signal intensities.

The presented results do not allow to draw conclusions on whether the performance of ML allows a more reliable diagnosis than that of a radiologist. In the future, prospective studies must therefore evaluate the clinical value of radiomics and investigate if it significantly improves radiological diagnosis.

Conclusion

In conclusion, ML using MRI-based radiomics allows sensitive and specific differentiation of benign and malignant and benign versus (pre-)malignant peripheral nerve sheath tumors in NF1 patients and represents a promising noninvasive diagnostic augmented radiological imaging tool. Feature expression distributions of premalignant atypical tumors illustrate biological plausibility of the considered radiomics characteristics. The presented techniques should be combined with established clinical and further image-based biomarkers and further validated in larger, prospective multi-center studies.

Keywords

atypical neurofibroma | machine learning | magnetic resonance imaging | malignant peripheral nerve sheath tumor| NF1

Funding

This study has been supported by a grant of the "Werner-Otto-Stiftung" (13/97) to L.W. and J.S. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest statement. The authors report no conflict of interest. All authors have approved the final version of the article.

Authorship statement. Study conception: I.R., J.S., L.W., R.W., F.M., V.F.M. Planning of imaging sequences: J.S., L.W. MRI data acquisition: I.R., J.S., L.W., I.M., F.W., F.S. Data and statistical analysis: I.R., F.M., R.W. Writing: I.R., R.W., F.M. Contribution in writing and revising the manuscript: J.S., L.W., G.A., P.B., S.F., F.W., F.S., I.M., V.F.M.

References

1. Lammert M, Friedman JM, Kluwe L, Mautner VF. Prevalence of neurofibromatosis 1 in German children at elementary school enrollment. *Arch Dermatol.* 2005;141(1):71–74.
2. Huson SM, Compston DA, Harper P. A genetic study of von Recklinghausen neurofibromatosis in south east Wales. II. Guidelines for genetic counselling. *J Med Genet.* 1989;26(11):712–721.
3. Fritchie K, Crago A, van de Rijn M. *Soft Tissue and Bone Tumours, WHO Classification of Tumours,* 5th ed., Vol. 3. Lyon, France: World Health Organization; 2020.
4. Tucker T, Wolkenstein P, Revuz J, Zeller J, Friedman JM. Association between benign and malignant peripheral nerve sheath tumors in NF1. *Neurology.* 2005;65(2):205–211.
5. Evans DGR, Baser ME, McGaughran J, et al. Malignant peripheral nerve sheath tumours in neurofibromatosis. *J Med Genet.* 2002;39(5):311–314.
6. Uusitalo E, Rantanen M, Kallionpää RA, et al. Distinctive cancer associations in patients with neurofibromatosis type 1. *J Clin Oncol.* 2016;34(17):1978–1986.
7. Farid M, Demicco EG, Garcia R, et al. Malignant peripheral nerve sheath tumors. *Oncologist.* 2014;19(2):193–201.
8. Miettinen MM, Antonescu CR, Fletcher CDM, et al. Histopathologic evaluation of atypical neurofibromatous tumors and their transformation into malignant peripheral nerve sheath tumor in patients with neurofibromatosis 1—a consensus overview. *Hum Pathol.* 2017;67:1–10.
9. Beert E, Brems H, Daniëls B, et al. Atypical neurofibromas in neurofibromatosis type 1 are premalignant tumors. *Genes Chromosom Cancer.* 2011;50(12):1021–1032.
10. Ahlawat S, Fayad LM, Khan MS, et al. Current whole-body MRI applications in the neurofibromatoses: NF1, NF2, and schwannomatosis. *Neurology.* 2016;87(7 Suppl 1):S31–S39.
11. Mautner VF, Asuagbor FA, Dombi E, et al. Assessment of benign tumor burden by whole-body MRI in patients with neurofibromatosis 1. *Neuro-Oncology.* 2008;10(4):593–598.
12. Tucker T, Friedman JM, Friedrich RE, et al. Longitudinal study of neurofibromatosis 1 associated plexiform neurofibromas. *J Med Genet.* 2009;46(2):81–85.
13. Nguyen R, Jett K, Harris GJ, et al. Benign whole body tumor volume is a risk factor for malignant peripheral nerve sheath tumors in neurofibromatosis type 1. *J Neurooncol.* 2014;116(2):307–313.
14. Wasa J, Nishida Y, Tsukushi S, et al. MRI features in the differentiation of malignant peripheral nerve sheath tumors and neurofibromas. 2012;194(6):1568–1574.
15. Matsumine A, Kusuzaki K, Nakamura T, et al. Differentiation between neurofibromas and malignant peripheral nerve sheath tumors in neurofibromatosis 1 evaluated by MRI. *J Cancer Res Clin Oncol.* 2009;135(7):891–900.
16. Demehri S, Belzberg A, Blakeley J, Fayad LM. Conventional and functional MR imaging of peripheral nerve sheath tumors: Initial experience. *Am J Neuroradiol.* 2014;35(8):1615–1620.
17. Well L, Salamon J, Kaul MG, et al. Differentiation of peripheral nerve sheath tumors in patients with neurofibromatosis type 1 using diffusion-weighted magnetic resonance imaging. *Neuro-Oncology.* 2019;21(4):508–516.
18. Derlin T, Tornquist K, Münster S, et al. Comparative effectiveness of ^{18}F-FDG PET/CT versus whole-body MRI for detection of malignant peripheral nerve sheath tumors in neurofibromatosis type 1. *Clin Nucl Med.* 2013;38(1):e19–e25.
19. Salamon J, Derlin T, Bannas P, et al. Evaluation of intratumoural heterogeneity on ^{18}F-FDG PET/CT for characterization of peripheral nerve

sheath tumours in neurofibromatosis type 1. *Eur J Nucl Med Mol Imaging.* 2013;40(5):685–692.
20. Salamon J, Veldhoen S, Apostolova I, et al. 18F-FDG PET/CT for detection of malignant peripheral nerve sheath tumours in neurofibromatosis type 1: tumour-to-liver ratio is superior to an SUV_{max} cut-off. *Eur Radiol.* 2014;24(2):405–412.
21. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* 2017;14(12):749–762.
22. 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.
23. Neurofibromatosis: Conference Statement. *Arch Neurol.* 1988;45(5):575–578.
24. Trojani M, Contesso G, Coindre JM, et al. Soft-tissue sarcomas of adults; study of pathological prognostic variables and definition of a histopathological grading system. *Int J Cancer.* 1984;33(1):37–42.
25. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage.* 2006;31(3):1116–1128.
26. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 2017;77(21):e104.
27. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res.* 2012;12:2825–2830.
28. Karpievitch YV, Hill EG, Leclerc AP, Dabney AR, Almeida JS. An introspective comparison of random forest-based classifiers for the analysis of cluster-correlated data by way of RF++. *PLoS One.* 2009;4(9):e7087.
29. Chalkidou A, O'Doherty MJ, Marsden PK. False discovery rates in PET and CT studies with texture features: a systematic review. *PLoS One.* 2015;10(5):e0124165.
30. Scott DW. *Multivariate Density Estimation.* John Wiley & Sons, Inc.; 1992.
31. Martin E, Geitenbeek RTJ, Coert JH, et al. A Bayesian approach for diagnostic accuracy of malignant peripheral nerve sheath tumors: a systematic review and meta-analysis. *Neuro-Oncology.* 2021;23(4):557–571.
32. Bernthal NM, Jones KB, Monument MJ, et al. Lost in translation: ambiguity in nerve sheath tumor nomenclature and its resultant treatment effect. *Cancers.* 2013;5(2):519–528.
33. 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.
34. Cook GJR, Lovat E, Siddique M, et al. Characterisation of malignant peripheral nerve sheath tumours in neurofibromatosis-1 using heterogeneity analysis of ^{18}F-FDG PET. *Eur J Nucl Med Mol Imaging.* 2017;44(11):1845–1852.
35. Gross AM, Wolters PL, Dombi E, et al. Selumetinib in children with inoperable plexiform neurofibromas. *N Engl J Med.* 2020;382(15):1430–1442.
36. Ahlawat S, Blakeley JO, Rodriguez FJ, Fayad LM. Imaging biomarkers for malignant peripheral nerve sheath tumors in neurofibromatosis type 1. *Neurology.* 2019;93(11):e1076–e1084.
37. 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 Central Eur Neurosurg.* 2016;77(6):495–504.
38. Matsumoto Y, Endo M, Harimaya K, et al. Malignant peripheral nerve sheath tumors presenting as spinal dumbbell tumors: clinical outcomes and characteristic imaging features. *Eur Spine J.* 2015;24(10):2119–2125.
39. Park JE, Park SY, Kim HJ, Kim HS. Reproducibility and generalizability in radiomics modeling: possible strategies in radiologic and statistical perspectives. *Korean J Radiol.* 2019;20(7):1124.
40. Well L, Dobel K, Kluwe L, et al. Genotype-phenotype correlation in neurofibromatosis type-1: NF1 whole gene deletions lead to high tumor-burden and increased tumor-growth. *PLoS Genet.* 2021;17(5):e1009517.
41. Liu Y, Jordan JT, Bredella MA, et al. Correlation between NF1 genotype and imaging phenotype on whole-body MRI. *Neurology.* 2020;94(24):e2521–e2531.
42. Baeßler B, Weiss K, dos Santos DP. Robustness and reproducibility of radiomics in magnetic resonance imaging: a phantom study. *Invest Radiol.* 2019;54(4):221–228.