Radiomics method in the differential diagnosis of diabetic foot osteomyelitis and charcot neuroarthropathy

Article *in* The British journal of radiology · April 2023

DOI: 10.1259/bjr.20220758

CITATIONS
4

READS
130

5 authors, including:

Ferhat Cuce
Gulhane Military Medical Academy
73 PUBLICATIONS **276** CITATIONS

SEE PROFILE

Kerim Bora Yilmaz
University of Health Sciences, Ankara Gulhane Education and Research HospitalTurkey
88 PUBLICATIONS **959** CITATIONS

SEE PROFILE

Title

Radiomics method in the differential diagnosis of diabetic foot osteomyelitis and Charcot neuroarthropathy

Manuscript Type: Musculoskeletal Imaging

Authors

Ferhat CUCE MD[1], Gokalp Tulum[2], Kerim Bora Yılmaz MD[3], Onur Osman[2], Ayse Aralasmak MD[4]

[1]Department of Radiology, Health Science University, Gulhane Training, and Research Hospital, Ankara, Turkey

[2] Department of Mechatronics Engineering, Engineering and Architecture Faculty, Nisantasi University, Istanbul, Turkey

[3] Department of General Surgery, Health Science University, Gulhane Training and Research Hospital, Ankara, Turkey

[4] Department of Radiology, Liv Hospital Vadi, Istanbul, Turkey

Corresponding Author:

Ferhat Cuce, MD

Ass. Prof. of Radiology

Department of Radiology, Health Science University, Gulhane Training and Research Hospital, 06018, Etlik, Ankara, Turkey

E-mail: ferhatcuce@gmail.com

Tel: +90 312 3044701

Funding: Each author has no grants or other assistance.

Author disclosures: Each author has no disclosures.

Abstract

Objectives: Our study used a radiomics method to differentiate bone marrow signal abnormality (BMSA) between Charcot neuroarthropathy (CN) and osteomyelitis (OM).

Materials and Method: The records of 166 patients with diabetic foot suspected CN or OM between January 2020 and March 2022 were retrospectively examined. A total of 41 patients with BMSA on MRI were included in this study. The diagnosis of OM was confirmed histologically in 24 of 41 patients. We clinically followed 17 patients as CN with laboratory tests. We also included 29 nondiabetic patients with traumatic (TR) BMSA on MRI as the third group. Contours of all BMSA on T1 and T2-weighted images in three patient groups were segmented semi-automatically on ManSeg (v.2.7d). The T1 and T2 features of three groups in radiomics were statistically evaluated. We applied multi-class classification (MCC) and binary-class classification (BCC) methodology to compare classification results.

Results: For MCC, the accuracy of Multi-Layer Perceptron (MLP) was 76.92% and 84.38% for T1 and T2, respectively. According to BCC, for CN, OM and TR BMSA, the sensitivity of MLP is 74%, 89.23%, and 76.19% for T1, and 90.57%, 85.92%, 86.81% for T2, respectively. For CN, OM, and TR BMSA, the specificity of MLP is 89.16%, 87.57%, and 90.72% for T1 and 93.55%, 89.94%, and 90.48% for T2 images, respectively.

Conclusion: In the diabetic foot, the radiomics method can differentiate the BMSA of CN and OM with high accuracy.

Advances in knowledge: The radiomics method can differentiate the BMSA of CN and OM with high accuracy.

Keywords: Charcot, neuroarthropathy, osteomyelitis, diabetic foot, radiomics

Radiomics method in the differential diagnosis of diabetic foot osteomyelitis and Charcot neuroarthropathy

Abstract

Objectives: Our study used a radiomics method to differentiate bone marrow signal abnormality (BMSA) between Charcot neuroarthropathy (CN) and osteomyelitis (OM).

Materials and Method: The records of 166 patients with diabetic foot suspected CN or OM between January 2020 and March 2022 were retrospectively examined. A total of 41 patients with BMSA on MRI were included in this study. The diagnosis of OM was confirmed histologically in 24 of 41 patients. We clinically followed 17 patients as CN with laboratory tests. We also included 29 nondiabetic patients with traumatic (TR) BMSA on MRI as the third group. Contours of all BMSA on T2 and T1 weighted images in three patient groups were segmented semi-automatically on ManSeg (v.2.7d). The T1 and T2 features of three groups in radiomics were statistically evaluated. We applied the multi-class classification (MCC) and binary-class classification (BCC) methodologies to compare results.

Results: For MCC, the accuracy of Multi-Layer Perceptron (MLP) was 76.92% and 84.38% for T1 and T2, respectively. According to BCC, for CN, OM and TR BMSA, the sensitivity of MLP is 74%, 89.23%, and 76.19% for T1, and 90.57%, 85.92%, 86.81% for T2, respectively. For CN, OM, and TR BMSA, the specificity of MLP is 89.16%, 87.57%, and 90.72% for T1 and 93.55%, 89.94%, and 90.48% for T2 images, respectively.

Conclusion: In diabetic foot, the radiomics method can differentiate the BMSA of CN and OM with high accuracy.

Advances in knowledge: The radiomics method can differentiate the BMSA of CN and OM with high accuracy.

Keywords: Charcot, neuroarthropathy, osteomyelitis, diabetic foot, radiomics

INTRODUCTION

Diabetes-related foot diseases are associated with high morbidity and substantial economic burdens worldwide. Diabetic patients have a 25% lifetime risk of developing foot ulcers, with the greatest risk for as many as 50% of the patients for subsequent osteomyelitis (OM) and amputation. [1] Five years mortality rate following below-knee amputation is 39% to 80% in patients with OM. [2,3] Early diagnosis and management of foot ulcers can avoid limb amputation.

The diabetic foot may present with Charcot neuroarthropathy (CN), OM, and infectious complications of soft tissue.

Diabetic polyneuropathy occurs in as many as 70% of patients and is the most common cause of foot osteoarthropathy.[4] Repetitive traumas based on sensory neuropathy and hyperemia due to autonomic neuropathy lead to osteoporosis and joint deformity.[5-7]

Dry skin due to autonomic neuropathy is sensitive to callus formation, and traumas break down the callus, contributing to skin ulceration.[8] The skin ulcer creates a portal for soft tissue infection and lays the groundwork for OM. Treatment of infection in diabetic feet is often problematic due to insufficient immune systems and hypoperfusion.

The clinical findings of CN may be difficult to distinguish from OM. In the acute phase of CN, the foot is characterized as erythematous, warm, and swollen. Hotfoot with no ulcer, acute phase of CN should primarily be considered with soft tissue infection or deep venous thrombus. [8, 9] Furthermore, CN and OM can co-exist as hotfoot with skin ulcers. However, the treatment strategies differ markedly; anti-biotherapy and surgery are the primary for OM, whereas protected weight-bearing is the mainstay for CN. [8]

Imaging plays a crucial role in distinguishing CN from OM and may guide early management on whether necessary to amputate. However, radiography has poor sensitivity and specificity in the differential of both entities; it is considered the first-line imaging investigation in diabetic hotfoot. After initial radiography, Magnetic Resonance Imaging (MRI) is the method of choice to diagnose OM, with a sensitivity of 90% and specificity of 79%. [11] Besides diagnosis, MRI, with its fine contrast resolution and anatomic detail, is well-suited to stage the extent of infection and the degree of tissue viability that is useful for guiding therapy .[12] Hyperintensity of bone marrow on T2-weighted images has a high sensitivity for OM but relatively low specificity unless a hypointensity accompanies it on T1-weighted images on MRI . [13, 14] Unfortunately, the marrow signal may sometimes be present similarly on both active CN and OM, [9, 15] and the location of the signal abnormality and soft tissue findings may be the only key features for differential diagnosis. OM occurs almost exclusively by the contiguous spread of infection to the bone from adjacent skin ulceration. [16] CN is not related to an overlying skin

ulcer and usually involves multiple midfoot bones and shows marrow abnormality in the periarticular and subchondral distribution.[10]

Although morphologic MR imaging is the most useful diagnostic method for diabetic foot, there are no clear distinguishing radiologic features between CN and OM. New functional MR imaging techniques derived from diffusion-weighted imaging (DWI) and dynamic contrast enhancement (DCE) can be combined with morphologic sequences to improve diagnostic accuracy. [11] Unfortunately, these functional sequences are not routinely used for diabetic foot assessment.

Radiomics is an advanced way to generate a high-dimensional feature set from radiologic images based on the distribution/relationship of image voxels and their statistical features. [17, 18] Features received from distribution and relationship patterns of voxels that incorporate the region of interest are primarily part of the radiomics statistics. Different transformation matrices of the radiological images, such as wavelet and curvelet transformation, can also be used to source feature sets. [19] As a source (input data set) of machine learning methods, radiomics can be carried out on different clinical decision-making factors and evaluation of response to treatment or diagnostic classifications.[20]

There has been no radiomics study on the differentiation of CN and OM in the diabetic foot. Few studies were published about the textural analysis reported in the literature on diabetic foot disease. [21,22] Our study aims to evaluate the potential of a machine learning algorithm via radiomics for differentiating the signal intensity of bone marrow between CN and OM.

MATERIALS AND METHODS

Patient

The local ethics committee approved this retrospective study. Written consent was waived.

The records of 166 patients with diabetic foot suspected CN or OM between January 2020 and March 2022 were retrospectively examined. A total of 41 diabetic patients who had bone marrow signal abnormality with hotfoot with/without skin ulcer were included in this study. They were scanned with the protocol of diabetic foot on the same Magnetic Resonance Imaging (MRI) machine. The diagnosis of OM was confirmed histologically in 24 of 41 patients. After excluding cellulitis and deep venous thrombosis, we clinically followed 17 patients as CN with laboratory tests. The diagnosis of CN was confirmed by the regression of clinical findings after offloading the extremity without antibiotic treatment. The study also included 29 nondiabetic patients with MRI with bone marrow signal abnormality after acute trauma (TR) as the third group compared with OM and CN groups. Eventually, the total number of patients was 70.

Imaging parameters

All MRIs were performed on a Philips 3T imaging system with a dedicated foot, ankle, and knee coil. All studies included fast spin-echo [FSE] T1-weighted (time echo [TE]: 6.6–20, repetition time [TR]: 400–646, echo train length [ETL]: 2–5), fat-saturated FSE T2-weighted (TE: 70–90, TR: 2,600–5,600, ETL: 10–12), and short tau inversion recovery imaging (STIR) imaging (TE: 30–70, TR: 2,900–4,500, ETL: 9–11, TI: 150–230, angle: 140). SPAIR T1-weighted fat-saturated imaging following IV gadolinium administration was reviewed when available.

Segmentation and Data Augmentation

For MRI image analysis, two independent radiologists determined the consensus area of the signal abnormality and semi-automated slice segmentation.

Firstly, the radiologists assessed the sagittal plane of T2 images. The readers selected contiguous images containing the signal abnormality of bone marrow near the skin ulcer and subarticular region. Secondly, the images corresponding to the area of abnormal signals detected in T2 were also recorded in sagittal images of T1. Both selected T2 and T1 images were saved as DICOM files and sent for segmentation. Post-contrast T1 images were not assessed. The DICOM images were uploaded to ManSeg (v.2.7d) software. Contours of all signal abnormality on T2 and T1 images were segmented semi-automatically on ManSeg (v.2.7d). In ManSeg, radiologists first delineated the ROI roughly, and then segmentation was finalized automatically with an active contour algorithm (Figure 1).

Sometimes the small size of the dataset may cause overfitting in classification. To avoid overfitting, we used different samples from different levels of the segments of the ROI as a new case, which is one of the data augmentation techniques performed in a radiomics-based machine learning study. [23] Eventually, for T1-weighted images, augmentation resulted in 299 labeled segmentation regions (64 CN, 137 OM, and 98 TR), and for T2-weighted images, augmentation resulted in 301 labeled segmentation regions (64 CN, 138 OM, and 99 TR) from 70 cases. Figure 2 shows the different samples of one OM case for T1 (A, B, and C) and T2 (D, E, and F) images respectively.

Feature Extraction

Before feature extraction, the ± 3σ method is preferred as a normalization technique for T1 and T2 weighted MRI images. [24] In this method, the intensity range of normalized images is converted between $min_{norm} = \mu - 3\sigma$ and $max_{norm} = \mu + 3\sigma$, where μ represents the mean and σ represents the standard deviation of the image intensities in the ROI. Radiomics features of the MRI datasets are extracted from original images, fine (kernel size of 3x3x1), medium (kernel size of 5x5x2), and coarse patterns (kernel size of 7x7x3) of Laplacian of Gaussian (LoG) filtered images and four different frequency sub-bands ((low-low, low-high, high-low and high-high) of wavelet decomposition of the images results. Due to the different morphological structures of the ROI, shape features of radiomics do not account for. The total number of

features is calculated at 736 per ROI. The description of the extracted features is given in Table 1.

Feature Selection

We proposed a two-layer cascade feature selection method for determining the optimal feature set. Firstly, the degree of collinearity is selected as a feature selector. Pearson's correlation coefficient (r) matrices of the features are calculated, and the r threshold is selected as 0.7. [25] The feature with the smallest p-value is selected as the first feature, and the features with low collinearity ($-0.7 \leq r \leq 0.7$) between the candidate feature and all previously selected features are included in the feature subset. As an output of the first layer of the feature selection method, the number of selected features is 47 and 48 for T1 and T2-weighted images, respectively. Secondly, the Neighborhood Component Feature Selection (NCFS) Algorithm [26] is applied, and features with feature weights smaller than 0.001 are filtered from the feature set. Here 0.001 is selected empirically. Figure 3 shows the weighted values of the NCFS algorithm. The final feature subsets are 5 and 9 for T1 and T2-weighted images, respectively.

Classification

To compare CN, TR, and OM classification results, we applied multi-class classification (MCC) and binary-class classification (BCC) methodologies for classification. For T1 and T2 weighted MRI cases, Multi-Layer Perceptron (MLP) and The Logistic Regression (LR) are selected as classifiers, and for training and evaluation, ten-fold cross-validation is used. For the MLP classifier, three different MLP structures (have two hidden layers, and the number of neurons in each hidden layer is generated randomly between the size of the input and double the size of the input) are constructed, and learning rate and momentum coefficient values of each structure are randomly determined between 0.2-0.6 and 0.5-0.9 respectively. The details of the structures of MLP and parameter settings are given in Supplementary material 1. For LR, multinomial logistic regression is constructed for MCC, and binomial logistic regression is constructed for BCC. The committee structure is decided on the final decision based on majority voting.

The average performance metrics across all 10 partitions are calculated for both T1 and T2 weighted MRI cases separately. The performance of MCC for two different classifiers is evaluated by confusion matrix and accuracy. The performance of BCC is evaluated by sensitivity, specificity, positive likelihood ratio, negative likelihood ratio, positive predictive value, negative predictive value, accuracy, and area under the receiver operating characteristics (AUC). Also, for BCC, receiver operating characteristic (ROC) curves are given in the Results.

RESULTS

Patients Demographics and Tumor Characteristics

In the OM group, 58.34% (n=14) of the patients were male and 41.66% (n=10) were female, and the mean age was 63.68±14.07 years. In the CN group, 52.95% (n=9) of the patients were male and 47.05% (n=8) were female, and the mean age was 61.13±11.56 years. In the TR group, 62.06% (n=18) of the patients were male and 37.94% (n=11) were female, and the mean age was 31.13±11.55 years.

Dimension Reduction (Feature Selection)

According to the degree of collinearity, the first layer of the feature selection algorithm reduced the number of features in the subset to 47 for T1-weighted and 48 for T2-weighted images among 736 features. Figure 4A and Figure 4B show the auto/cross-correlation matrix of the T1 and T2 weighted image features, respectively. As expected from the first layer of the feature selection method, there is no significant collinearity ($|r|<0.7$) among the features. The second layer of the feature selection algorithm has calculated the weights of the selected feature by using a diagonal adaptation of neighborhood component analysis (NCA). [27] Eventually, the proposed feature selection algorithm selected features that had feature weights smaller than 0.001. Nine features are selected as the final feature subset for classification for T1 weighted images, five features, and T1 weighted images.

When the final feature subset is investigated, for T1-weighted images, two features are derived from original images (feature1 and feature2). The rest are derived from LoG filters with different sizes of kernels (feature3, feature four, and feature5). For T2-weighted images, while two features are generated from the original image (feature1 and feature2), features three and four originated from the LoG filter with kernel size 7x7x3, and the rest of the features are generated from different frequency bands of the wavelet transform. The detailed description of the final feature subsets of T1 and T2 weighted images is given in Table 2, and their boxplot graphs are shown in Figure 5A for T1-weighted images and Figure 5B for T2-weighted images.

Classification

When the two classifiers' accuracy is compared, MLP values are better than LR, with 76.92% accuracy for T1-weighted images and 84.38% for T2-weighted images. Detailed metrics are given in Table 3 and Table 4. Also, for BCC, MLP performance is better statistical values than LR. For CN, the sensitivity of MLP is 74% and 90.57% for T1 and T2 weighted images,

respectively. For MLP, sensitivity values of OM are calculated at 89.23% and 85.92%, and sensitivity values of TR are calculated at 76.19% and 86.81%, respectively, for T1 and T2 weighted images. Specificity values of CN, OM, and TR are 89.16%, 87.57%, and 90.72% for T1-weighted cases, respectively. Specificity values of T2-weighted cases for MLP are obtained at 93.55%, 89.94%, and 90.48% for CN, OM, and TR, respectively. Detailed metrics are given in Table 5 for T1 and T2 weighted images.

According to the results, T2-weighted images have better classification performance than T1-weighted images for both MCC and BCC.

DISCUSSION

The diabetic foot may present with CN, OM, and soft-tissue complications, including cellulitis, myositis, ulceration, sinus tracts, and abscess. Morphologic MR imaging is the most useful diagnostic method for diabetic foot; however, there are no clear distinguishing radiologic features between CN and OM. [9, 15] Moreover, the treatment strategies differ markedly; antibiotics and surgical debridement are the primaries for infection, whereas protected weight-bearing is the mainstay for CN. [8] The location of the signal abnormality and soft tissue findings may be the only key features for differential diagnosis. OM occurs almost exclusively by the contiguous spread of infection to the bone from adjacent skin ulceration. [16] CN is not related to an overlying skin ulcer and usually involves multiple midfoot bones and shows marrow abnormality in the periarticular and subchondral distribution. [10] We evaluated the potential of a machine learning algorithm via radiomics for differentiating the signal intensity of bone marrow among CN, OM, and TR cases. T1 and T2 images of cases were taken into consideration.

When final feature subsets of T1-weighted images are investigated, all features are derived from the original (feature one and feature 2) and the LoG filter of MRI images with different kernel sizes (feature 3, feature four, and feature 5). Also, two features (feature one and feature 4) are the same from different image types, derived as a contrast. This roughly may mean that spatial intensity changes have the main role in generating features for T1-weighted images. Because LoG filters are used to detect areas of rapid changes in images after the noise reduction and contrast is a measure of intensity changes between voxels and their neighborhood. Figure 5A shows that the contrast feature of the original image (feature 1) has the highest mean value while the contrast feature of the LoG filter (feature 4) has the smallest value, which can mean that the local intensity variation of ROI can be a selective feature before and after the noise reduction and sharpening procedures.

When the final feature subset of T2-weighted images is examined, it can be seen that features are generated from the original and both wavelet and LoG transformation of the MRI images. But the main source of the feature subset is wavelet transforms of the MRI, and skewness which measures the asymmetry of the distribution about the mean, is the dominant feature (feature 2, feature 3, and feature 5). The details and the visualized images of the wavelet transformations and LoG filter are given in Supplementary material 2. Also, the definition of each transformation matrix and the visualized images of each matrix are added to supplementary material 2. The difference in the asymmetry of the distribution about to the mean for the T1 and T2-weighted image is shown in Figure 1 and Figure 2 given in Supplementary File 2.

Wavelet analysis is a decomposition method to divide information on an image into different components. Wavelet transform of a grayscale image is passed through high-pass and low-pass filters, and the image is decomposed into high and low-frequency components at every level we get 4 sub-signals. When Table II (final feature subset) investigated, wavelet decomposition is not only represented key information but also transformation matrices such as GLCM (Gray Level Co-Occurrence Matrix), GLSZM (Gray Level Size Zone Matrix) and NGTDM (Neighboring Gray Tone Difference Matrix) have represented key information about to determine final feature final subset.

For wavelet decomposition, it can be said that high and low-frequency components of the decomposition levels can highlight the importance of T2-weighted images.

In this study firstly, each class is classified separately (MCC). Secondly, the classification performance of BCC is calculated with two different classifiers where one class is selected as positive while the remaining classes are negative.

According to the MCC results, MLP has the highest classification ratio with an accuracy of 76.92% and 84.38% for T1 and T2-weighted images, respectively. For BCC, MLP also performs better for T1-weighted (AUC values of CN, OM, and TR are 0.818, 0.896, 0.918, respectively) and T2-weighted images (AUC values of CN, OM, and TR are 0.93, 0.909, 0.898 respectively).

When T1 and T2-weighted images are compared, T2-weighted images are better at imagining modalities for classifying the CN, OM, and TR. According to the ROC graphics of the MLP for T1-weighted MRI (Figure 6) classification performance of the TR and the OM, cases have nearly the same and better AUC values than in CN cases. For T2-weighted MRI, three cases' classification performances (Figure 7) are approximately the same AUC values.

When Table 3 is investigated, the confusion chart of MLP shows that architecture distinguished each class of diseases better than LR for both T1 and T2-weighted images. For T1-weighted images, 16 of the CN cases misclassified as TR for MLP and the highest classification error occurred for CN cases for LR. For T2-weighted images, MLP classified CN

cases correctly according to T1-weighted cases but the LR algorithm again produced the highest classification error for CN cases. 19 CN cases were classified as OM and 16 cases of CN were classified as TR.

According to the BCC and MCC, when the performance of radiomics features is investigated, while classification performance for each strategy has a reasonable accuracy (Table 4 and Table 5), the classification performance of BCC is better than the MCC. The selected radiomics features best distinguished the CN from other diseases on T2-weighted images with an accuracy of 93.02%.

To our knowledge, this is the first study that classifies CN, OM, and TR by using radiomics features. Therefore, we could not compare our results with the literature. There are few studies on the diagnosis of non-tumoral pathological signals in the bone marrow with the radiomics technique.[28]

Our study had some limitations. First, it is a retrospective study since diabetic foot OM is not very common in daily MRI practice. The second limitation is the low number of patients in our single-center study. Due to the low number of cases, we had to apply feature selection and classification steps to all data and this may lead to a bias. Therefore, it is expected that the performance of the classifier will be adversely affected by unseen new data. However, obtaining MRI images on the same machine and with the same examination protocol is one of the strengths of our study in terms of data homogeneity.

REFERENCES

1. Lipsky BA. Osteomyelitis of the foot in diabetic patients. Clin Infect Dis 1997;25(6):1318–1326.
2. Boulton AJM, Vileikyte L, Ragnarson-Tennvall G, Apelqvist J. The global burden of diabetic foot disease. Lancet 2005;366(9498):1719–1724.
3. Singh N, Armstrong DG, Lipsky BA. Preventing foot ulcers in patients with diabetes. JAMA 2005;293(2):217–228.
4. Gouveri E, Papanas N. Charcot osteoarthropathy in diabetes: a brief review with an emphasis on clinical practice. World J Diabetes 2011; 2: 59-65.
5. Russell JM, Peterson JJ, Bancroft LW. MR imaging of the diabetic foot. Magn Reson Imaging Clin N Am 2008;16(1):59-70.
6. Chatha DS, Cunningham PM, Schweitzer ME. MR imaging of the diabetic foot: diagnostic challenges. Radiol Clin North Am 2005;43(4):747–759.
7. Jones EA, Manaster BJ, May DA, Disler DG. Neuropathic osteoarthropathy: diagnostic dilemmas and differential diagnosis. RadioGraphics 2000;20(Spec Issue):279–293.

8. Jonathan CB, Jennifer LD, Nicholas GR, Daniel EW, David AR. Diabetic musculoskeletal complications and their imaging mimics. Radiographics 2012;32(7):1959-74.
9. Fatma Bilge E, Saziye Eser S, Ali O. Charcot foot in diabetes and an update on imaging. Diabet Foot Ankle 2013;20(4):124-127.
10. Sunil Y, Ankit T. A Pictorial Review of Diabetic foot Manifestations. Med J Malaysia 2013;68: 279-289.
11. Martín NT, Luna AA, Beltrán LS, Gómez CM, et al. Advanced MR Imaging Techniques for Differentiation of Neuropathic Arthropathy and Osteomyelitis in the Diabetic Foot. Radiographics. 2017;37(4):1161-1180.
12. Donovan A, Schweitzer ME. Current concepts in imaging diabetic pedal osteomyelitis. Radiol Clin North Am 2008;46(6):1105–1124.
13. Collins MS, Schaar MM, Wenger DE, Mandrekar JN. T1-weighted MRI characteristics of pedal osteomyelitis. AJR Am J Roentgenol 2005;185(2):386–393.
14. Johnson PW, Collins MS, Wenger DE. Diagnostic utility of T1-weighted MRI characteristics in the evaluation of osteomyelitis of the foot. AJR Am J Roentgenol 2009;192(1):96–100.
15. Rogers LC, Bevilacqua NJ. Imaging of the Charcot's foot. Clin Podiatr Med Surg 2008; 25: 263–74.
16. Ledermann HP, Morrison WB, Schweitzer ME. MR image analysis of pedal osteomyelitis: distribution, patterns of spread, and frequency of associated ulceration and septic arthritis. Radiology 2002;223(3):747–755.
17. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology 2016;278:563–577.
18. Timmeren JE, Caster D, Tanadini-Lang S, et al. Radiomics in medical imaging—"how-to"guide and critical reflection. Insights Imaging 2020;11:91-92.
19. Jing R, Wang J, Li J, et al. A wavelet features derived radiomics nomogram for prediction of malignant and benign early‐stage lung nodules. Sci. Rep. 2021;11: 223-30.
20. Jeong J, Yin Y, Wangg H, et al. A review of original articles published in the emerging field of radiomics. Eur. J. Radiol. 2020;127: 89-91.
21. Stavros S, Samar A, Qiong Q, et al. Machine learning prediction of diabetic foot ulcers in the inpatient population. Vascular. 2021;30:145-152
22. Xie P, Li Y, Deng B, Du C, Rui S, et al. An explainable machine learning model for predicting in-hospital amputation rate of patients with diabetic foot ulcer. Int Wound J. 2022;19(4):910-918.
23. Kocak B, Durmaz ES, Kaya O, et al. Machine learning-based unenhanced CT texture analysis for predicting BAP1 mutation status of clear cell renal cell carcinomas. Acta Radiol. 2020;61(6):856-864.
24. Collet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. Magn Reson Imaging 2004;22:81–91.
25. Dormann CF, Elith J, Bacher S, et al. Collinearity: A review of methods to deal with it

and a simulation study evaluating their performance. Ecography 2013;36(1):27-46
26. Yang W., Wang K., Zuo W. Neighborhood component feature selection for high-dimensional data. J. Comput.2012;7(1):161-168.
27. Goldberger J., Hinton G., Roweis S., Salakhutdinov R. Neighbourhood Components Analysis. Advances in Neural Information Processing Systems. 2004;17:513-520.
28. Michail E. Klontzas, Georgios C, et al. Radiomics and Machine Learning Can Differentiate Transient Osteoporosis from Avascular Necrosis of the Hip. Diagnostics (Basel) 2021 Sep; 11(9): 1686

Figure 1 — Semi-automated segmentation procedure by using ManSeg(v.2.7b)

Click here to access/download;Figure;Figure_1.jpg

Figure 2 — Examples for augmentation techniques: different samples of one OM case for T1 (A, B and C) and T2 (D, E and F) images respectively.

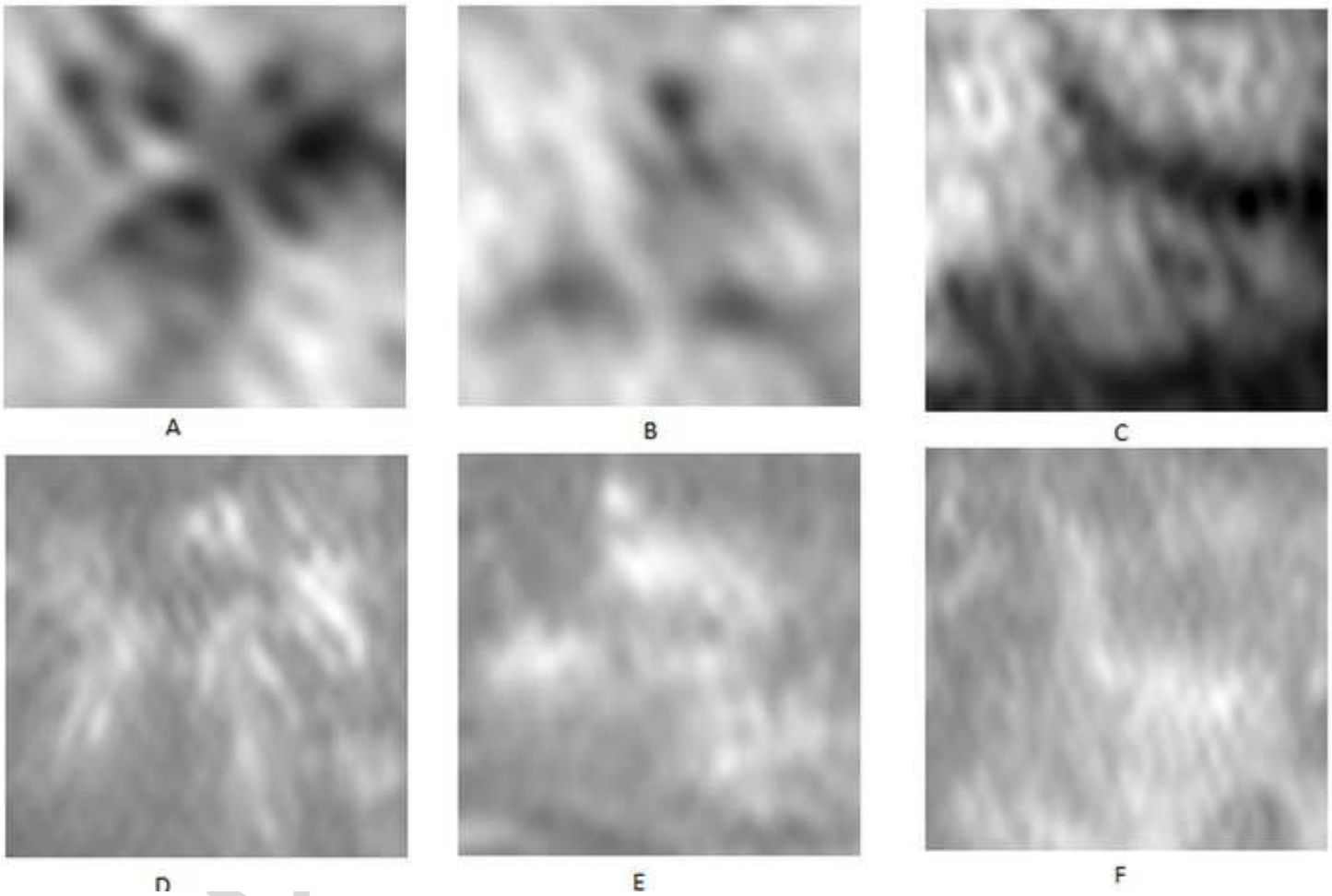


Figure 3 — Feature weights of the NCFS algorithm

Click here to access/download;Figure;Figure_3.jpg

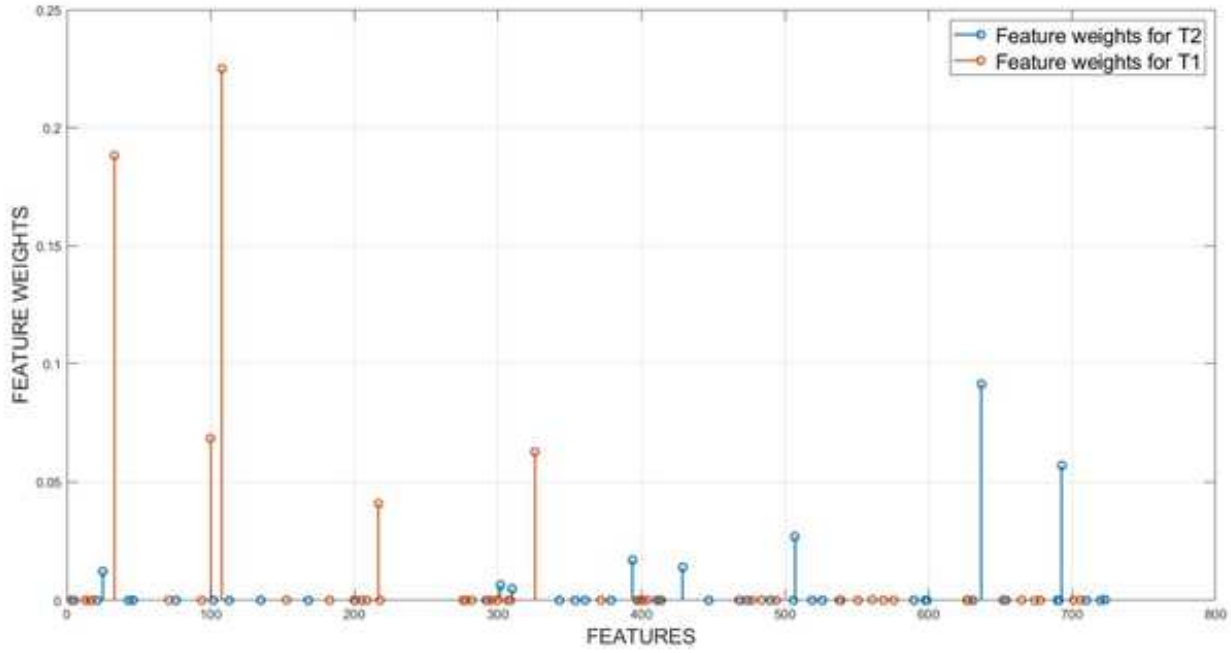


Figure 4 — The auto/cross-correlation matrix of the features for T1 and T2-weighted Images

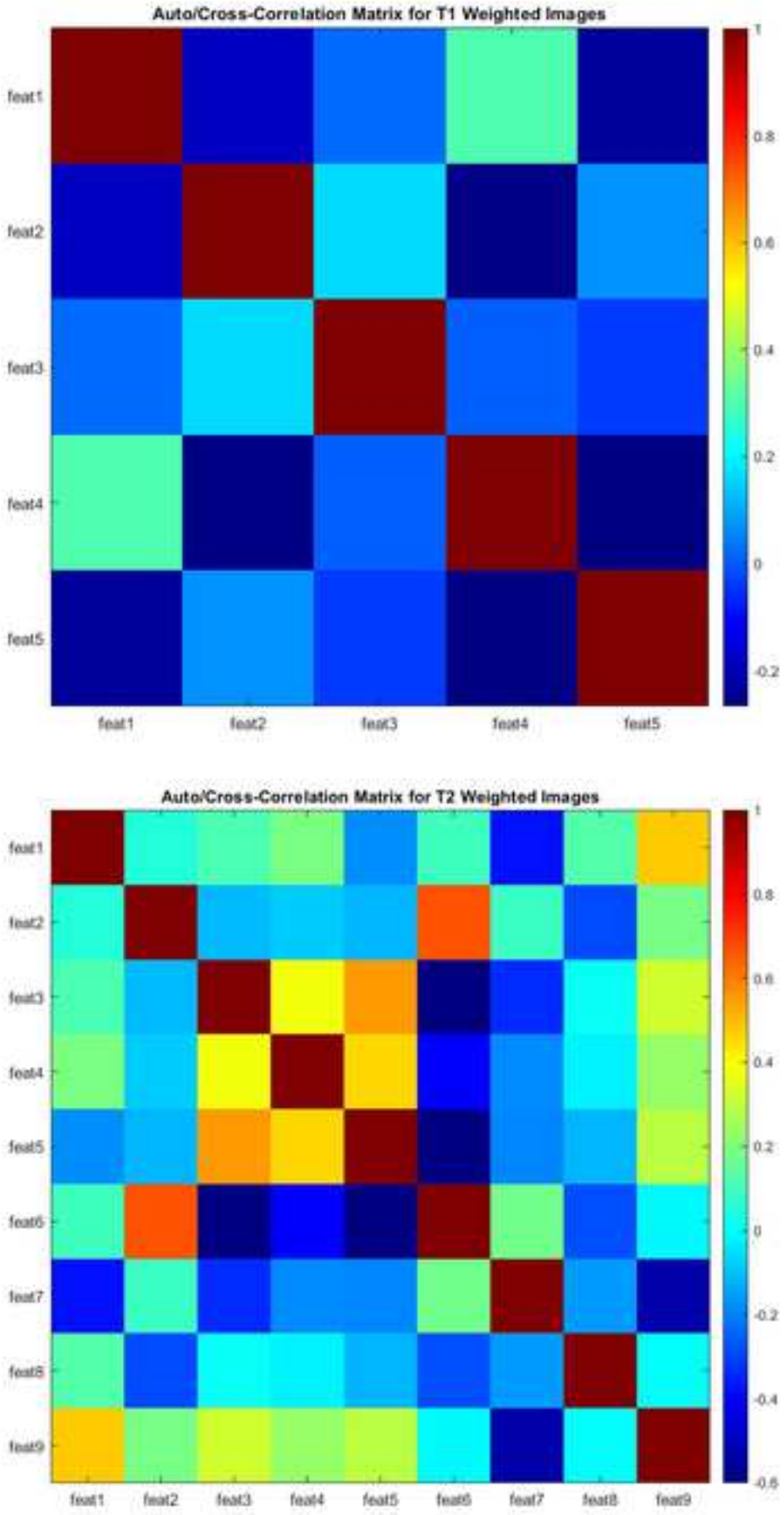


Figure 5 — Boxplot graphs of features for T1 and T2-weighted images

Click here to access/download;Figure;Figure_5.jpg

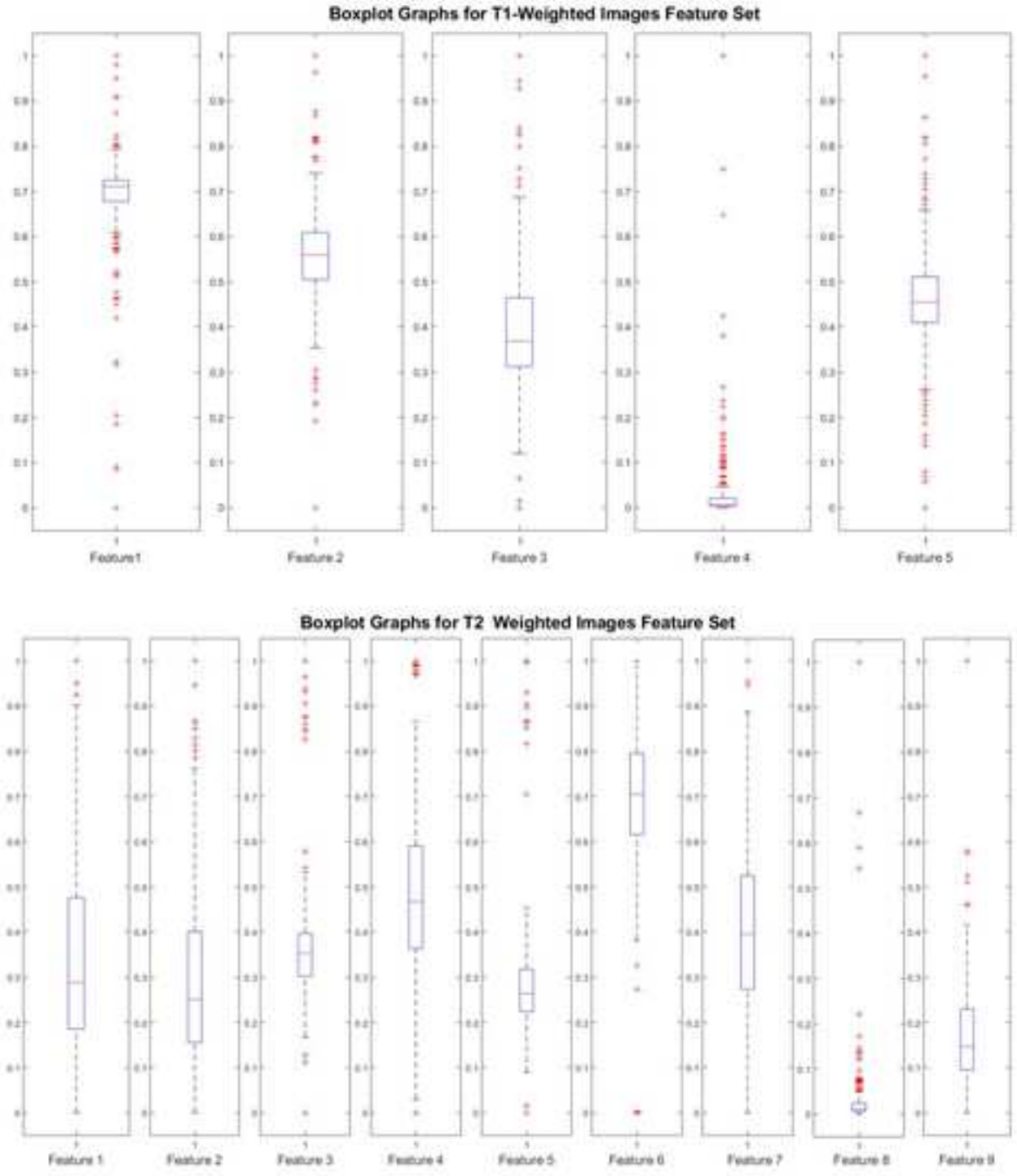


Figure 6 — ROC curves of OvA for T1-Weighted Images (MLP and LR)

Click here to access/download;Figure;Figure_6.jpg

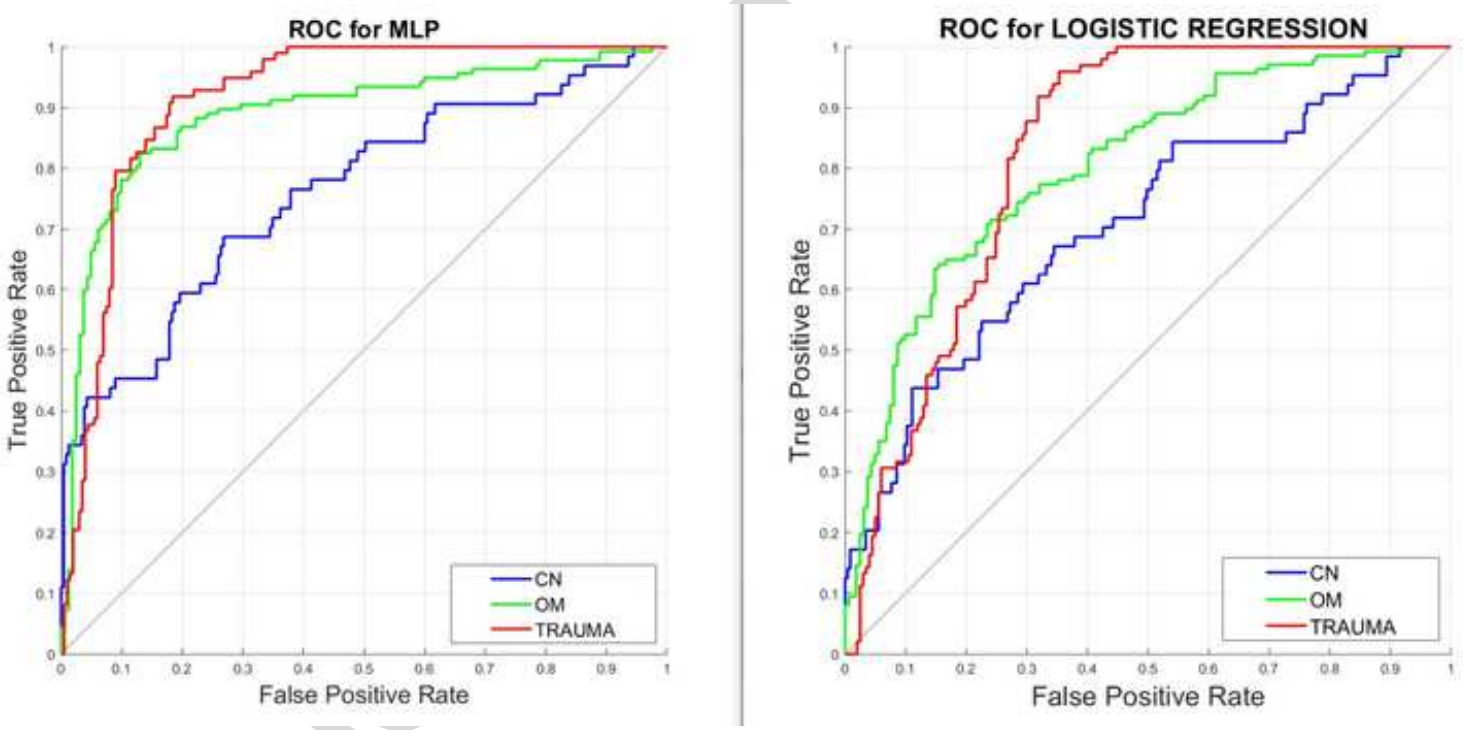
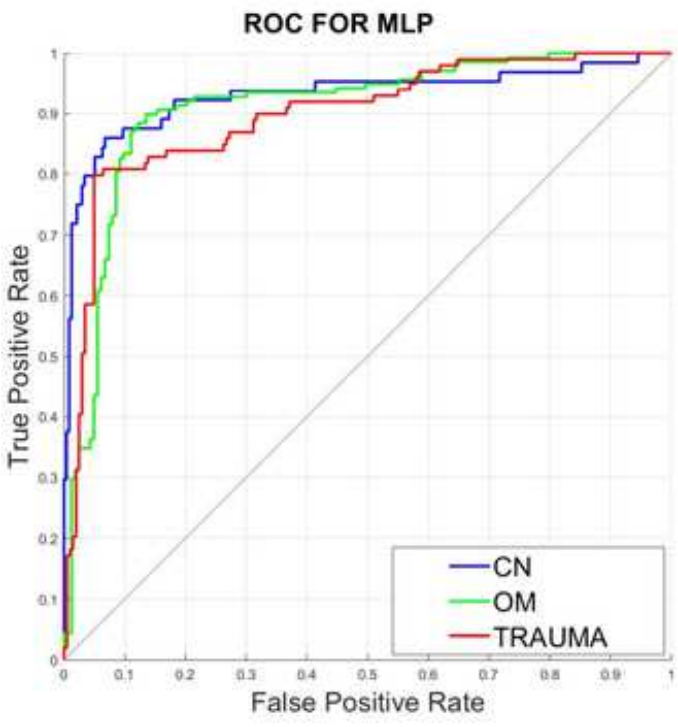


Figure 7 — ROC curves of OvA for T2-Weighted Images (MLP and LR)

Click here to access/download;Figure;Figure_7.jpg



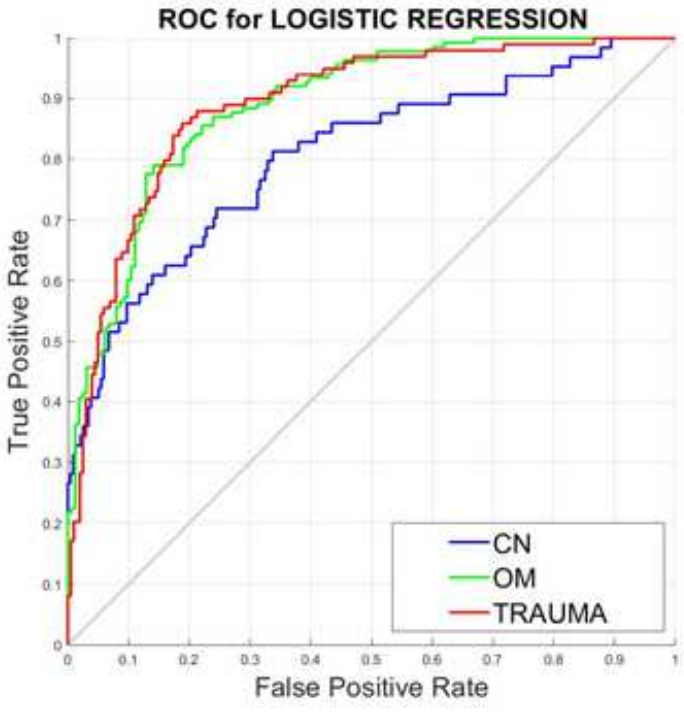


TABLE 1: Description of the extracted features

		NUMBER OF FEATURES	TOTAL NUMBER OF FEATURES
ORIGINAL IMAGE	1. FIRST ORDER STATISTICS	17	92
	2. GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM) FEATURES	24	
	3. GRAY LEVEL SIZE ZONE MATRIX (GLSZM) FEATURES	16	
	4. GRAY LEVEL RUN LENGTH MATRIX (GLRLM) FEATURES	16	
	5. NEIGHBOURING GRAY TONE DIFFERENCE MATRIX (NGTDM) FEATURES	5	
	6. GRAY LEVEL DEPENDENCE MATRIX (GLDM) FEATURES	14	
LoG FILTER (FINE, MEDIUM, COARSE PATTERNS)	1. FIRST ORDER STATISTICS	51	276
	2. GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM) FEATURES	72	
	3. GRAY LEVEL SIZE ZONE MATRIX (GLSZM) FEATURES	48	
	4. GRAY LEVEL RUN LENGTH MATRIX (GLRLM) FEATURES	48	
	5. NEIGHBOURING GRAY TONE DIFFERENCE MATRIX (NGTDM) FEATURES	15	
	6. GRAY LEVEL DEPENDENCE MATRIX (GLDM) FEATURES	42	
WAVELET TRANSFORM (LL-LH-HL-HH)	1. FIRST ORDER STATISTICS	68	368
	2. GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM) FEATURES	96	
	3. GRAY LEVEL SIZE ZONE MATRIX (GLSZM) FEATURES	64	
	4. GRAY LEVEL RUN LENGTH MATRIX (GLRLM) FEATURES	64	
	5. NEIGHBOURING GRAY TONE DIFFERENCE MATRIX (NGTDM) FEATURES	20	
	6. GRAY LEVEL DEPENDENCE MATRIX (GLDM) FEATURES	56	

TABLE 2: Description of the final feature subsets for T1 and T2-weighted images

T1-WEIGHTED IMAGES FINAL FEATURE SUBSET			
CODE	**FEATURE NAME**	**FEATURE CLASS**	**IMAGE TYPE**
f1	Contrast	GLCM	ORIGINAL IMAGE
f2	SDHGLE	GLDM	ORIGINAL IMAGE
f3	90th percentile	FIRST ORDER STATISTICS	LoG (3x3x1)
f4	Contrast	GLCM	LoG (5x5x2)
f5	Sum Entropy	GLCM	LoG (7x7x3)
T2-WEIGHTED IMAGES FINAL FEATURE SUBSET			
CODE	**FEATURE NAME**	**FEATURE CLASS**	**IMAGE TYPE**
f1	10th percentile	FIRST ORDER STATISTICS	ORIGINAL IMAGE
f2	Skewness	FIRST ORDER STATISTICS	ORIGINAL IMAGE
f3	Skewness	FIRST ORDER STATISTICS	LoG (7x7x3)
f4	Correlation	GLCM	LoG (7x7x3)
f5	Skewness	FIRST ORDER STATISTICS	WAVELET LL
f6	Zone Entropy (ZE)	GLSZM	WAVELET LL
f7	Inverse Variance	GLCM	WAVELET LH
f8	Contrast	NGTDM	WAVELET HL
f9	Sum Average	GLCM	WAVELET HH

* f: feature, GLCM: Gray Level Co-Occurrence Matrix, SDHGLE: Small Dependence High Gray Level Emphasis, GLDM: Gray Level Dependence Matrix, GLSZM: Gray Level Size Zone Matrix, NGTDM: Neighboring Gray Tone Difference Matrix, LL: low-low frequency band, HL: high-low frequency band, LH: low-high frequency band, 3x3x1/5x5x2/7x7x3: kernel size of the LoG filter

TABLE 3: Confusion matrix of OvO for both T1 and T2 weighted images

T1-WEIGHTED IMAGES							
OvO	MLP			OvO	LR		
	CN	OM	TR		CN	OM	TR
CN	39	9	16	CN	14	22	28
OM	9	118	10	OM	5	104	28
TR	6	6	86	TR	3	16	79
T2-WEIGHTED IMAGES							
OvO	MLP			OvO	LR		
	CH	OM	TR		CN	OM	TR
CN	50	10	4	CN	29	19	16
OM	5	124	9	OM	5	117	16
TR	4	15	80	TR	3	13	83

TABLE 4: Accuracy of OvO for both T1 and T2 weighted images

T1-WEIGHTED IMAGES			**T1-WEIGHTED IMAGES**		
OvO	MLP	LR	OvO	MLP	LR
Accuracy	76.92%	65.88%	Accuracy	84.38%	76.08%

Table 5: Performance metrics of T1/T2-weighted images (OvA) for both algorithms

T1- WEIGHTED IMAGES												
	MLP						LR					
Statistic	CN		OM		TR		CN		OM		TR	
	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI
Sensitivity	**74.00%**	59.66% - 85.37%	**89.23%**	82.59% - 93.99%	**76.19%**	66.89% - 83.96%	**68.75%**	41.34% - 88.98%	**77.88%**	69.10% - 85.14%	**58.00%**	47.71% - 67.80%
Specificity	**89.16%**	84.62% - 92.73%	**87.57%**	81.63% - 92.14%	**90.72%**	85.73% - 94.41%	**81.27%**	76.23% - 85.65%	**73.66%**	66.71% - 79.83%	**79.90%**	73.65% - 85.23%
Positive Likelihood Ratio	**6,82**	4.61 - 10.10	**7,18**	4.79 - 10.76	**8,21**	5.22 - 12.91	**3,67**	2.44 - 5.53	**2,96**	2.28 - 3.83	**2,89**	2.09 - 3.99
Negative Likelihood Ratio	**0.29**	0.18 - 0.47	**0.12**	0.07 - 0.20	**0.26**	0.19 - 0.37	**0.38**	0.19 - 0.80	**0.30**	0.21 - 0.43	**0.53**	0.41 - 0.67
Disease prevalence	**16.72%**	12.67% - 21.44%	**43.48%**	37.78% - 49.31%	**35.12%**	29.71% - 40.82%	**5.35%**	3.09% - 8.54%	**37.79%**	32.27% - 43.55%	**33.44%**	28.12% - 39.10%
Positive Predictive Value	**57.81%**	48.07% - 66.98%	**84.67%**	78.66% - 89.22%	**81.63%**	73.86% - 87.48%	**17.19%**	12.11% - 23.82%	**64.23%**	58.08% - 69.95%	**59.18%**	51.21% - 66.71%
Negative Predictive Value	**94.47%**	91.44% - 96.47%	**91.36%**	86.53% - 94.56%	**87.56%**	83.29% - 90.86%	**97.87%**	95.69% - 98.96%	**84.57%**	79.33% - 88.67%	**79.10%**	74.85% - 82.81%
Accuracy	**86.62%**	82.23% - 90.27%	**88.29%**	84.10% - 91.71%	**85.62%**	81.12% - 89.39%	**80.60%**	75.66% - 84.93%	**75.25%**	69.96% - 80.04%	**72.58%**	67.14% - 77.55%
AUC	**0,818**	0.774 - 0.862	**0,896**	0.861 - 0.931	**0,918**	0.887 - 0.949	**0.707**	0.655 - 0.759	**0.804**	0.759 - 0.849	**0.827**	0.784 - 0.870

T2- WEIGHTED IMAGES												
	MLP						LR					
Statistic	CN		OM		TR		CN		OM		TR	
	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI
Sensitivity	**90.57%**	79.34% - 96.87%	**85.92%**	79.09% - 91.18%	**86.81%**	78.10% - 93.00%	**78.57%**	59.05% - 91.70%	**80.74%**	73.07% - 87.02%	**77.53%**	65.72% - 84.19%
Specificity	**93.55%**	89.73% - 96.27%	**89.94%**	84.17% - 94.14%	**90.48%**	85.67% - 94.09%	**84.62%**	79.78% - 88.68%	**82.53%**	75.88% - 87.98%	**85.71%**	80.24% - 90.15%
Positive Likelihood Ratio	**14,4**	8.67 - 22.73	**8,54**	5.34 - 13.65	**9,12**	5.96 - 13.94	**5,11**	3.64 - 7.17	**4,62**	3.29 - 6.50	**5,43**	3.74 - 7.54
Negative Likelihood Ratio	**0.10**	0.04 - 0.23	**0.16**	0.10 - 0.24	**0.15**	0.09 - 0.25	**0.25**	0.12 - 0.52	**0.23**	0.16 - 0.33	**0.26**	0.20 - 0.41
Disease prevalence	**17.61%**	13.48% - 22.39%	**47.18%**	41.42% - 52.99%	**30.23%**	25.10% - 35.76%	**9.30%**	6.27% - 13.16%	**44.85%**	39.14% - 50.66%	**29.77%**	25.10% - 35.76%
Positive Predictive Value	**75.00%**	64.95% - 82.93%	**88.41%**	82.66% - 92.42%	**79.80%**	72.10% - 85.79%	**34.37%**	27.18% - 42.37%	**78.99%**	72.78% - 84.09%	**69.70%**	61.82% - 76.57%
Negative Predictive Value	**97.89%**	95.27% - 99.07%	**87.73%**	82.60% - 91.50%	**94.06%**	90.32% - 96.41%	**97.47%**	94.98% - 98.74%	**84.05%**	78.74% - 88.23%	**90.00%**	84.99% - 92.20%

Accuracy	**93.02%**	89.53% - 95.63%	**88.04%**	83.83% - 91.48%	**89.37%**	85.32% - 92.61%	**84.05%**	79.42% - 88.00%	**81.73%**	76.89% - 85.93%	**83.28%**	77.97% - 86.82%
AUC	**0,93**	0.901- 0.959	**0,909**	0.876- 0.941	**0,898**	0.864 - 0.932	**0,805**	0.760 - 0.850	**0,891**	0.856 - 0.926	**0,89**	0.855 - 0.925