Differential Diagnosis of Diabetic Foot Osteomyelitis and Charcot Neuropathic Osteoarthropathy with Deep Learning Methods

Maide Cakir[1] · Gökalp Tulum[2] · Ferhat Cuce[3] · Kerim Bora Yilmaz[4] · Ayse Aralasmak[5] · Muhammet İkbal Isik[6] · Hüseyin Canbolat[7]

Abstract

Our study aims to evaluate the potential of a deep learning (DL) algorithm for differentiating the signal intensity of bone marrow between osteomyelitis (OM), Charcot neuropathic osteoarthropathy (CNO), and trauma (TR). The local ethics committee approved this retrospective study. From 148 patients, segmentation resulted in 679 labeled regions for T1-weighted images (comprising 151 CNO, 257 OM, and 271 TR) and 714 labeled regions for T2-weighted images (consisting of 160 CNO, 272 OM, and 282 TR). We employed both multi-class classification (MCC) and binary-class classification (BCC) approaches to compare the classification outcomes of CNO, TR, and OM. The ResNet-50 and the EfficientNet-b0 accuracy values were computed at 96.2% and 97.1%, respectively, for T1-weighted images. Additionally, accuracy values for ResNet-50 and the EfficientNet-b0 were determined at 95.6% and 96.8%, respectively, for T2-weighted images. Also, according to BCC for CNO, OM, and TR, the sensitivity of ResNet-50 is 91.1%, 92.4%, and 96.6% and the sensitivity of EfficientNet-b0 is 93.2%, 97.6%, and 98.1% for T1, respectively. For CNO, OM, and TR, the sensitivity of ResNet-50 is 94.9%, 83.6%, and 97.9% and the sensitivity of EfficientNet-b is 95.6%, 85.2%, and 98.6% for T2, respectively. The specificity values of ResNet-50 for CNO, OM, and TR in T1-weighted images are 98.1%, 97.9%, and 94.7% and 98.6%, 97.5%, and 96.7% in T2-weighted images respectively. Similarly, for EfficientNet-b0, the specificity values are 98.9%, 98.7%, and 98.4% and 99.1%, 98.5%, and 98.7% for T1-weighted and T2-weighted images respectively. In the diabetic foot, deep learning methods serve as a non-invasive tool to differentiate CNO, OM, and TR with high accuracy.

Keywords Osteomyelitis · Charcot neuropathic osteoarthropathy · Deep learning · Diabetic foot

✉ Maide Cakir
  maidecakrr@gmail.com

1 Department of Electrical Engineering, Faculty of Engineering and Natural Sciences, Bandirma Onyedi Eylul University, Balikesir, Turkey

2 Department of Electrical and Electronics Engineering, Istanbul Topkapi University, Engineering Faculty, Istanbul, Turkey

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

4 Department of General Surgery, Health Science University, Gulhane Training and Research, Ankara, Turkey

5 Department of Radiology, Liv Hospital Vadi, Istanbul, Turkey

6 Department of Radiology, Health Sciences University, Gulhane Training and Research Hospital, Ankara, Turkey

7 Department of Electrical and Electronics Engineering, Faculty of Engineering and Natural Sciences, Ankara Yildirim Beyazit University, Ankara, Turkey

Introduction

Diabetic foot osteomyelitis (OM) and Charcot neuropathic osteoarthropathy (CNO) are two common complications that can occur in patients with diabetes mellitus. While both conditions can present with similar symptoms, such as redness, swelling, and pain in the foot, they have different underlying causes and require different treatment approaches [1]. Imaging modalities help differentiate between the two conditions and guide treatment decisions. Magnetic resonance imaging (MRI) is useful for evaluating soft tissue and bone marrow changes and can help differentiate between OM and CNO [2]. In CNO, MRI can show bone marrow edema, joint effusions, and ligamentous injuries. The marrow edema is typically seen in the subarticular region adjacent to the affected joint [3]. MRI can show soft tissue defects and inflammation in OM and signal abnormality in the adjacent bone marrow. The bone marrow signal abnormality (BMSA)

is almost seen near the soft tissue infection than in CNO [4]. Due to the neuropathy, patients may also stub their feet. Post-traumatic bone marrow edema is another condition that can be confused with OM and CNO. Similar signal abnormalities of bone marrow in visual assessment such as T1 iso-hypointensity and T2 hyperintensity are observed in OM, CNO, and TR. While MRI can be a helpful tool in differentiating between OM and CNO, there are no clear distinguishing MRI features between the two conditions using morphologic sequences alone. However, new functional MRI techniques, derived from diffusion-weighted imaging (DWI) and dynamic contrast enhancement (DCE), can be combined with morphologic sequences to improve diagnostic accuracy [5]. By combining morphologic MRI sequences with DWI and DCE-MRI, it may be possible to improve the diagnosis. Unfortunately, these advanced imaging techniques are not standardized or routinely used. However, in bridging this gap, deep learning models can serve as valuable inputs for decision-making processes and diagnostic classifications. Therefore, in this study, we presented a non-invasive tool for differentiating bone marrow signal abnormalities using deep learning methods.

In recent years, computer vision and machine learning algorithms have been used in the prognosis and diagnosis of diseases across a wide range of medical imaging modalities. These modalities include ultrasound, computer tomography (CT), MRI, and whole slide imaging (WSI) [6, 7]. Deep learning (DL)-based convolutional neural networks (CNN) have advanced to the forefront of medical segmentation and classification because of their ability to discover and interpret patterns [8]. DL techniques are used by researchers in the field of diabetic foot ulcer (DFU) assessment for detection and recognition [9, 10]. DFUQUTNET, DFUNET, ComparisonNet, EfficientNet, and ResNet are among the deep learning methods proposed for DFU identification, with most designed for visual images [11]. Diez et al. compared the diagnostic accuracy of dynamic contrast–enhanced magnetic resonance imaging (DCE-MRI) and DWI with two regions of interest (ROI) sizes with 18-fluoro-deoxyglucose positron emission tomography/computed tomography (18-FDG PET/CT) for distinguishing OM from CNO [12]. The research has constraints, such as its exclusive focus on a single institution, a small sample size, technical difficulties, exclusion of patients with renal failure—a common complication of diabetes—dependence on ROI-based analysis, the influence of vascular diseases on the estimation of arterial input function when calculating DCE-MRI parameters, and a failure to investigate additional MRI parameters. Yap et al. summarize the outcomes of DFUC2020 [9, 13] by evaluating the DL-based algorithms such as Faster R–CNN, its three variations, and an ensemble technique, Efficient-Det, YOLOv3, YOLOv5, and a novel Cascade Attention Network [14]. Due to the limited number of datasets, some improvement methods remained at a minimum level. Goyal et al. presented the refined EfficientDet algorithms for DFU detection in the DFUC dataset [15], which contains 4500 images of ulcerated feet [16]. This study utilized the DFUC2020 dataset obtained from a particular hospital, potentially lacking a comprehensive representation of the diverse characteristics of DFUs across various populations or clinical environments. Additionally, the research concentrated on the technical facets of deep learning architecture for detecting DFUs, without delving into a thorough discussion of the clinical validation and real-world implementation of the proposed method. Munadi et al. developed a new framework for DFU classification combining thermal imaging, decision fusion, and deep neural networks [11]. The study is constrained by its reliance on a sole public dataset, potentially lacking a precise representation of the entire population. Hernandez-Guedes et al. proposed a novel strategy to evaluate the effectiveness of DL models in situations where labeled data were scarce [17]. As a case study, this methodology was examined by putting into practice many different experimental configurations to classify early-stage DFU samples taken from a plantar thermogram dataset. In this research, the scarcity of annotated data in deep learning models, especially within the healthcare domain, presents difficulties in evaluating performance metrics and detecting overfitting. This can potentially result in biased outcomes and make it challenging to gauge the representativeness of the model. Muralidhara et al. suggested a novel CNN for discriminating between five diabetes mellitus (DM) and non-DM severity grades from plantar thermal images and evaluated its performance to that of pre-trained networks like AlexNet [18]. The study is constrained by the difficulty of implementing standardized thermal imaging methods consistently across diverse acquisition systems. This obstacle impedes the generalizability and practical application of decision support systems based on thermography. Ahsan et al. implemented DL frameworks, AlexNet, GoogLeNet, VGG16/19, ResNet50.101, SqueezeNet, MobileNet, and DenseNet for the classification of ischemia and infection using the benchmark dataset DFU2020 [19]. The study has limitations such as data imbalance and a limited amount of data. Alzubaidi et al. present a novel dataset of 754-ft images containing diabetic ulcer skin and healthy skin from several individuals [20]. DFU QUTNet, a unique Deep-CNN, is proposed for automatically categorizing abnormal skin (DFU) and normal skin (healthy skin). The study suggests that DFUNet is presently the most advanced network for classifying DFUs. Nonetheless, it would be advantageous to conduct a more comprehensive comparison with alternative techniques and networks to present a more convincing demonstration of its superiority. Thotad et al. implemented EfficientNet for 844 diabetic foot ulcer images for early prognosis and diagnosis of DFUs [21]. The study is constrained by

its capability to only classify a foot as either diabetic (abnormal) or healthy (normal), lacking the capacity to evaluate the severity of the condition or furnish detailed information about complications. Anaya-Isaza and Zequera-Diaz investigated different ML methods of using foot thermography to identify diabetic patients [22]. The primary objective of the study was to identify a suitable classification index, such as the Thermal Change Index (TCI). The study acknowledges the restricted collection of thermographic image data, potentially impacting the applicability of the findings to a more extensive population.

In the literature, there are also relevant approaches for distinguishing bone marrow alterations in MRI. Chuah et al. [23] explored the potential use of MRI-based texture features for classifying individuals with and without bone marrow lesions. The study, involving 58 subjects, identified 29 with bone marrow lesions. Texture features were computed for the distal femur's weight-bearing region, and a forward feature selection method identified a subset for classification. The results revealed that 98 out of 147 features exhibited statistically significant differences between normal and affected marrows. Subject classification achieved an AUC of 0.914, and slice classification obtained an AUC of 0.780. Limitations were the exclusive focus on osteoarthritis-related bone marrow lesions and the high number of false positives and negatives. Li et al. [24] investigated the predictive capacity of MRI-based three-dimensional texture analysis on ınfrapatellar fat pad (IPFP) abnormalities for incident radiographic knee osteoarthritis in 690 at-risk participants. AUC values for the clinical score, ınfrapatellar fat pad texture score, and MRI Osteoarthritis Knee Score in the test cohort were 0.65, 0.84, and 0.85, respectively. Two main limitations included concerns about result validity due to the lack of independence between the test and development cohorts and the inability to conduct histopathologic examinations, introducing uncertainty about the IPFP texture's association with histopathologic characteristics. Kostopoulos et al. [25] aimed to identify textural changes in knee lesions (BME, INJ, OST) using 121 MRI knee examinations. Cases were grouped based on radiological findings. The study achieved an AUC of 0.93 ± 0.02 in the test set using combined radiomic descriptors. A limitation lies in relying on data division into groups based on radiological findings.

In the literature, the researchers mainly focused on the classification of the DFU and distinguishing bone marrow alterations. To the best of our knowledge, there is no study published in the English literature about the use of DL for the classification of OM and CNO in diabetic foot. We also included non-diabetic patients with traumatic BMSA in our study. Although the same abnormal signals can be seen in the bone marrow in all three clinics (OM, CNO, TR), there have been different histopathological background mechanisms. We added the TR group to strengthen our diagnostic accuracy. Our study aims to evaluate the potential of a DL algorithm for differentiating the signal intensity of bone marrow among OM, CNO, and TR in diabetic and non-diabetic patients.

Materials and Methods

Materials

Patient

The Health Sciences University Scientific Research Ethics Committee approved the study Ethics Committee approved the study (date: 16.05.2023, decision number: 2023-199).

The medical records of patients exhibiting suspected diabetic foot complications on foot MRI between September 2016 and December 2022 were reviewed, and the local ethics committee approved this retrospective study. Written consent was waived.

One hundred seventeen patients were scanned on the same MRI machine with the diabetic foot protocol. Patients with MRIs with poor diagnostic quality were excluded. Finally, eighty-seven diabetic patients who had BMSA on MRI with hotfoot with/without skin ulcers were included. Sixty-four diabetic patients with BMSA on MRI were diagnosed as OM histopathologically. The CNO group consisted of 23 patients considered clinically and laboratory as Charcot and responded to related conservative treatments during their follow-up. Sixty-one non-diabetic patients with BMSA on foot MRI after acute TR were also included as the third group named TR. The total number of patients in all three groups (OM, CNO, and TR) was 148.

In the OM group, 68.8% ($n = 44$) of the patients were male, 31.2% ($n = 20$) were female, and the mean age was 59 ± 16.58. In the CNO group, 60.9% ($n = 14$) of the patients were male, 39.1% ($n = 9$) were female, and the mean age was 53.22 ± 16.52 years. In the TR group, 67.2% ($n = 41$) of the patients were male, 32.8% ($n = 20$) were female, and the mean age was 43.22 ± 16.52 years.

Imaging Parameters

All MRIs were performed on a Philips 3 T imaging system with a dedicated foot and ankle coil. The studies consisted of fast spin-echo (FSE) T1-weighted imaging [time echo (TE) 6.6–25, repetition time (RT) 400–700, echo train length (ETL) 2–6], fat-saturated FSE T2-weighted imaging [TE 60–90, RT 2600–6600, ETL 9–13], and short tau inversion recovery (STIR) imaging [TE 30–80, RT 2,900–4800, ETL 9–13, TI 150–250, angle 140]. SPAIR T1-weighted fat-saturated imaging following IV gadolinium administration was reviewed when available.

2D Image Segmentation

The two independent musculoskeletal radiologists with more than 10 years of experience assessed the MRIs. In the visual evaluation of T1 and T2, iso-hypointensity of T1 and hyperintensity of T2 were seen in the bone marrow, which could not distinguish three different clinics (OM, CNO, and TR) from each other. For 2D image analysis, upon consensus, the radiologists segmented the area of the BMSA semiautomatically using ManSeg (v.2.7 g) software [26]. First, the radiologists determined MRI slices containing the signal abnormality of bone marrow near the skin ulcer and subarticular region for T1- and T2-weighted images. Then, the DICOM images were uploaded to ManSeg, and radiologists roughly delineated the contours of the BMSAs manually. Finally, segmentation was automatically done according to

the radiologist's drawings with an active contour algorithm. In Fig. 1, the process of delineating the OM ROI area for T1-weighted images is illustrated. Figures related to the ROI areas obtained through the ManSeg process for OM, CNO, and TR for both T1- and T2-weighted images are provided in the Supplementary File.

To prevent possible overfitting (caused by the limited dataset), we augmented sample space by selecting abnormal signals bigger than half of the largest cross-sectional area of the BMSA in one MRI image as a new sample case. Eventually, for T1-weighted images, 679 labeled regions (151 CNO, 257 OM, and 271 TR) and, for T2-weighted images, 714 labeled regions (160 CNO, 272 OM, and 282 TR) were segmented from 148 patients. Figures 2 and 3 show three different samples of OM (A–C), CNO (D–F), and TR (G–I) for T1- and T2-weighted images, respectively.
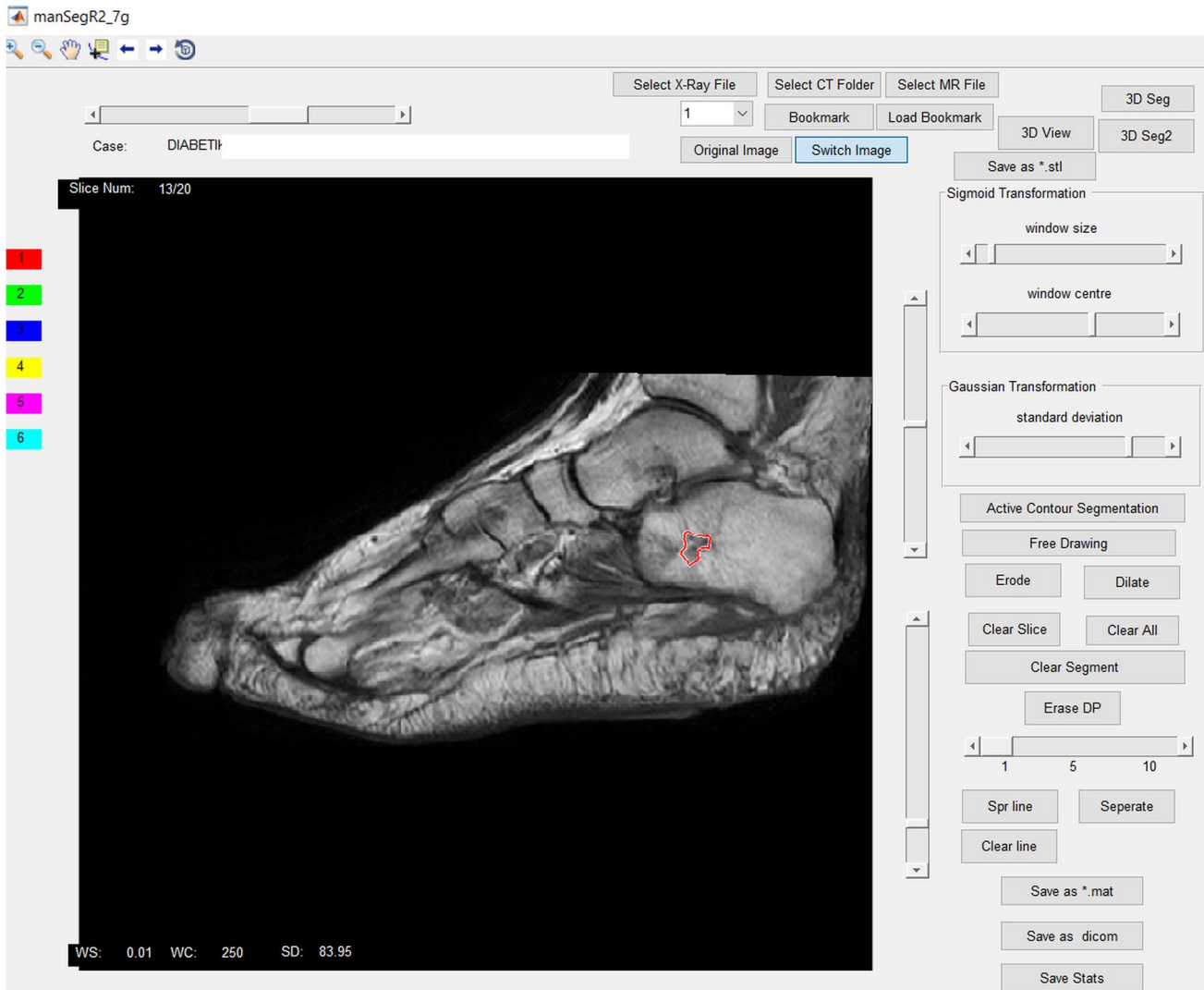

Fig. 1 The semi-automated segmentation process for identifying osteomyelitis (OM) region of interest (ROI) in T1-weighted image was conducted using ManSeg (v.2.7 g)
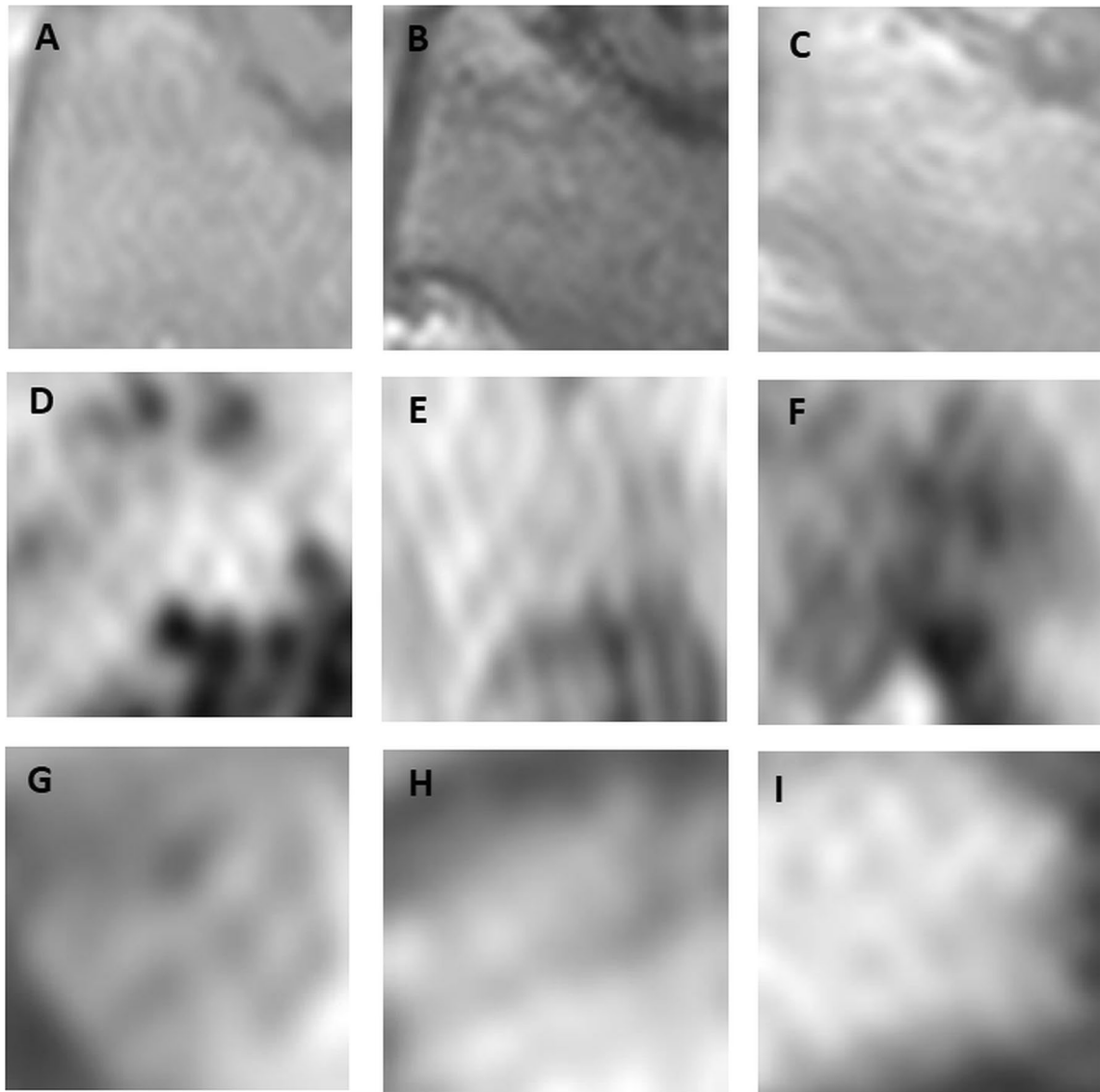
Fig. 2 Three different samples of OM (**A–C**), CNO (**D–F**), and TR (**G–I**) for T1-weighted images respectively

Methods

Data Preprocessing and Data Augmentation

The ROIs of the BMSA were extracted according to the radiologist's semi-automated drawings on ManSeg (v.2.7 g), and the bounding box of the BMSA was extracted as input images. ROIs were normalized using the Z-score normalization, and the normalized values of the voxels were transformed into grayscale values. The size of each image was interpolated to 224×224. Conversion of grayscale images to RGB was done as R=G=B = grayscale value. For data augmentation, rotation (randomly between $[-15^0, 15^0]$), reflection in the x- and y-axes, scaling (randomly between $[0.9, 1.1]$), and translation in the x- and y-axes (randomly between $[-10, 10]$ pixels) were performed. Finally, for T1-weighted images, 755 samples for CNO, 1285 samples for OM, and 1355 samples for TR and, for T2-weighted images, 800 samples for CNO, 1360 samples for OM, and 1410 samples for TR were generated. In the Supplementary File, we incorporated augmentation samples for OM, CNO, and TR in both T1- and T2-weighted images, totaling three distinct variations.

Classification

We performed randomly five different train-test cross-validation splits, but for each run, we employed the same threefold cross-validation for the algorithms, yielding 15 performance metric values for each algorithm. The average
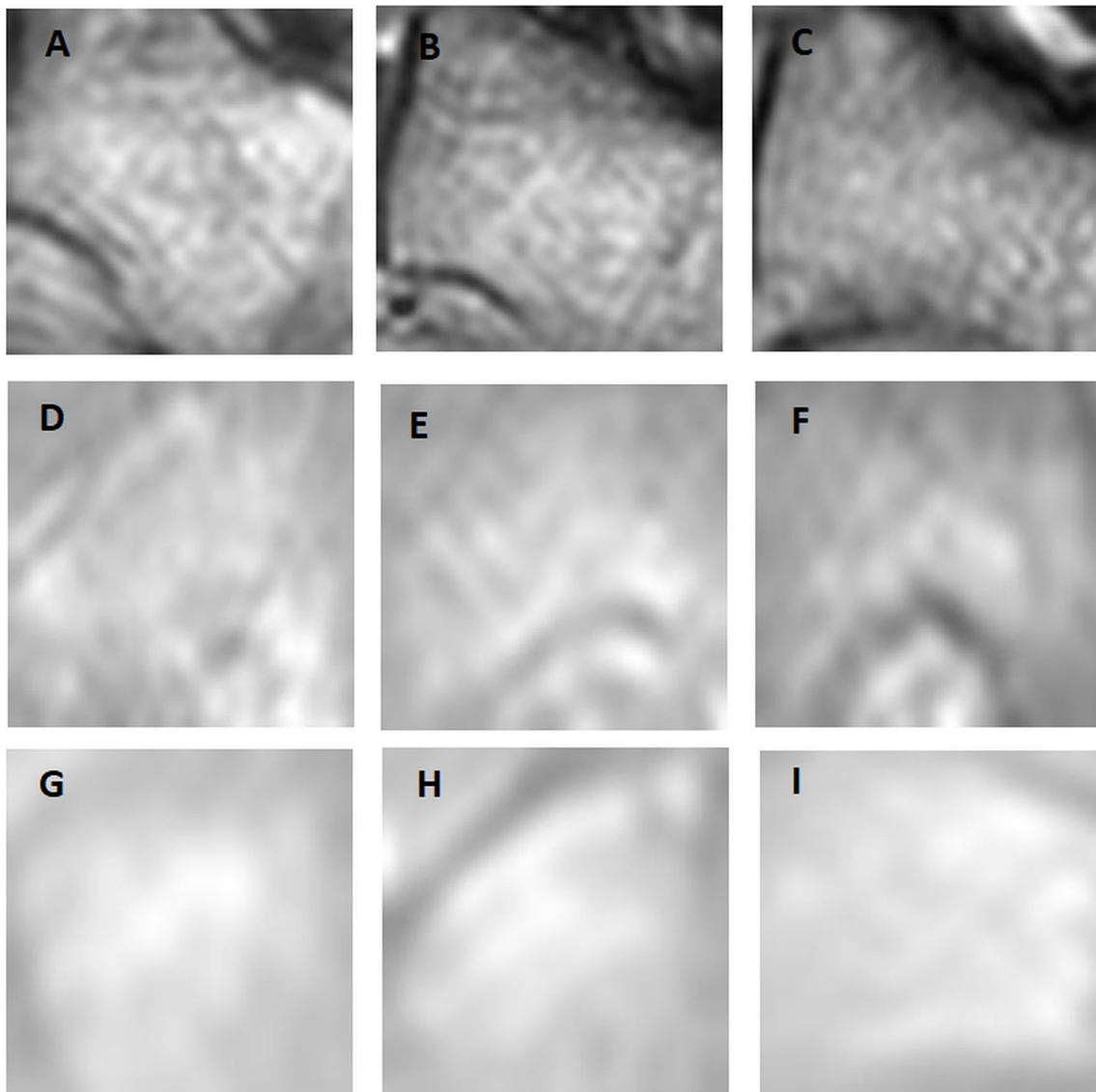
Fig. 3 Three different samples of OM (**A–C**), CNO (**D–F**), and TR (**G–I**) for T2-weighted images respectively

validation accuracy was given as a classification accuracy for both T1- and T2-weighted MRI cases. The flowchart of the study is provided in Fig. 4. We also employed similar split approaches (conducting five distinct random train-test cross-validation splits while utilizing consistent threefold cross-validation in each iteration) for binary-class classification (BCC), opting for a one-vs.-rest (OvR) strategy to compare the classification outcomes of CNO, TR, and OM. ResNet-50 [27] and EfficientNet-b0 [28] CNN models were used as a DL classifier, and training was evaluated using a batch size 32. Cross-entropy was selected as a loss function for both EfficientNet-b0 and ResNet-50. The learning rate and the number of epochs were selected as 0.0001 and 100, respectively. The Adam optimizer was utilized for training. Due to the small size of the input, the networks were trained

by applying transfer learning, using EfficientNet-b0 and ResNet-50 pre-trained weights. The training process was performed in Matlab 2021a. For training, a GPU-enabled (NVIDIA GeForce GTX 1060, 6 GB) system (Intel(R) Core (TM) i7-8700 CPU @ 3.20 GHz 3.19 GHz, 16,0 GB RAM) was used. The.m files and classification results were publicly available at [29].

Results

Table 1 shows that the ResNet-50 and the EfficientNet-b0 mean accuracy values were computed at 96.17% and 97.11%, respectively, for T1-weighted images. Additionally, accuracy values for ResNet-50 and EfficientNet-b0 were

Fig. 4 The flowchart of the proposed model
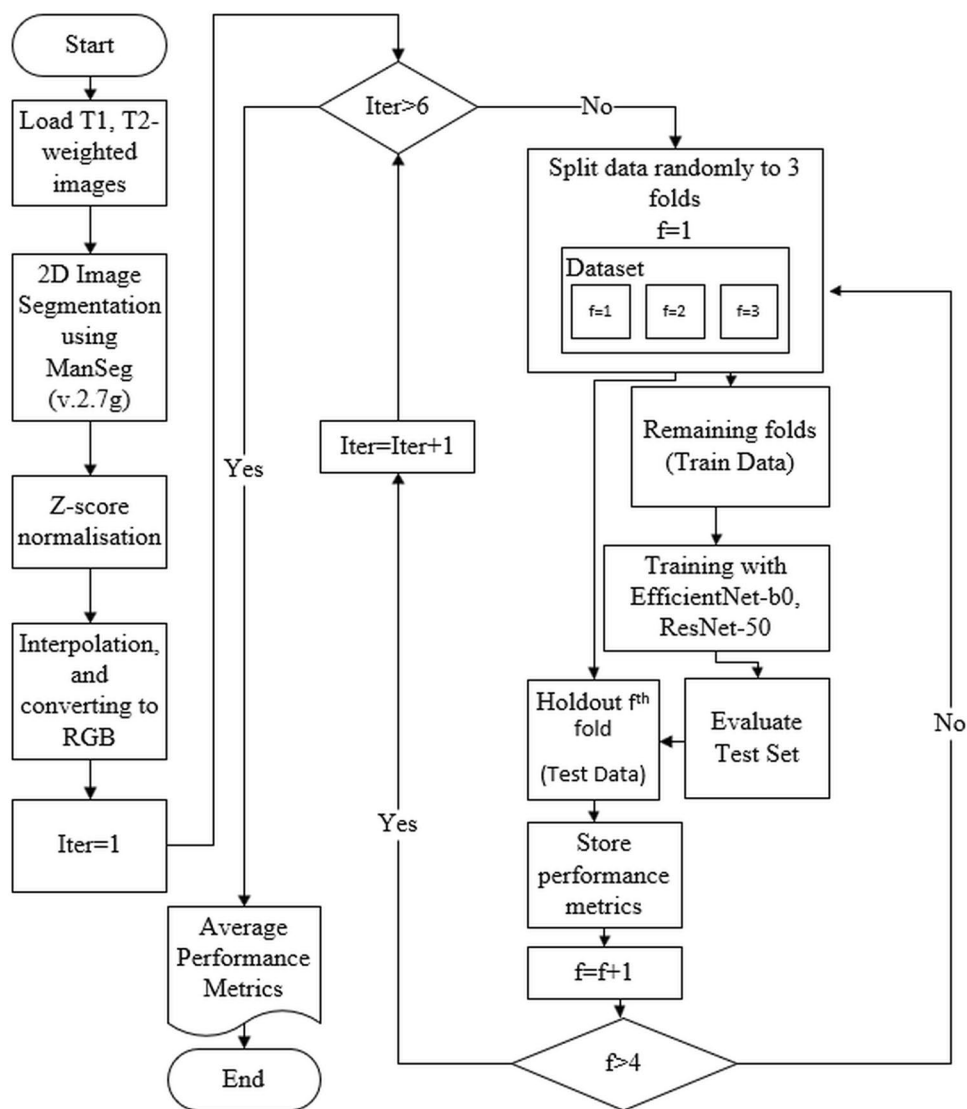
determined at 95.57% and 96.83%, respectively, for T2-weighted images. The ResNet-50 model's training for T1- and T2-weighted images concluded after 6 h 50 min 35 s and 6 h 47 min 25 s over 100 epochs, while the training of the EfficientNet-b0 model for the same T1- and T2-weighted images was finalized in 13 h 17 min 19 s and 13 h 29 min 36 s over 100 epochs, respectively. Training loss graphics per one-fold of cross-validation are provided in the Supplementary File. Upon examining the inference speeds, it was noted that ResNet-50 demonstrated faster processing, completing inference in approximately 5.56 s for T1-weighted images and 5.82 s for T2-weighted images. In contrast, EfficientNet-b0 exhibited slightly slower speeds, with inference times of around 7.93 s for T1-weighted images and 8.49 s for T2-weighted images. When the accuracy values of classifiers were examined for the same weighted images and different MRI sequences (T1 and T2), both ResNet-50 and EfficientNet-b0 demonstrated the

capability to effectively distinguish multi-class classification problems while working with datasets of small to medium sizes. According to the results of the t-tests, the observed accuracy metrics were statistically significant ($p < 0.05$). The overall mean value of confusion matrices for 15 test results of ResNet-50 and the EfficientNet-b0 was given in Table 1.

The OvR method has been preferred for BCC. This enables the calculation of mean sensitivity, specificity, and F1 scores for each class, providing clear insights into the performance of individual classes. For BCC, ResNet-50 demonstrates a sensitivity of 91.13% for CNO in T1-weighted images and 94.88% in T2-weighted images. ResNet-50 achieves 92.37% and 83.60% sensitivity rates for OM and 96.62% and 97.96% for TR in T1- and T2-weighted images, respectively. The specificity values for CNO, OM, and TR in T1-weighted cases are 98.07%, 97.96%, and 94.69%, respectively. Similarly, for T2-weighted cases, the specificity values for CNO, OM, and TR with ResNet-50 are 98.56%,

Table 1 The mean value of confusion matrices, (rounded to the nearest) and mean accuracy of classifiers for T1- and T2-weighted images

Confusion Matrix for ResNet-50

		T1-WEIGHTED IMAGES			T2-WEIGHTED IMAGES		
		Predicted Class			Predicted Class		
		CNO	OM	TR	CNO	OM	TR
Target Class	CNO	694	46	15	721	35	44
	OM	6	1266	13	17	1316	27
	TR	15	35	1305	9	26	1375
Accuracy (%)		96.17			95.57		
95% CI (%)		95.52–96.82			94.9–96.24		

Confusion Matrix for EfficientNet-b0

		T1-WEIGHTED IMAGES			T2-WEIGHTED IMAGES		
		Predicted Class			Predicted Class		
		CNO	OM	TR	CNO	OM	TR
Target Class	CNO	728	19	8	755	25	20
	OM	20	1245	20	20	1327	13
	TR	16	15	1324	9	26	1375
Accuracy (%)		97.11			96.83		
95% CI (%)		96.55–97.67			96.26–97.4		

CNO Charcot neuropathic osteoarthropathy, *OM* osteomyelitis, *TR* trauma

Table 2 The performance metrics for both algorithms in binary-class classification (BCC) using a one-vs.-rest (OvR) approach on T1/T2-weighted images

T1-weighted images

Statistic	ResNet-50						EfficientNet-b0					
	CNO		OM		TR		CNO		OM		TR	
	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI
Sensitivity	91.13%	90.17–92.09	92.37%	91.48–93.26	96.62%	96.01–97.23	93.17%	92.32–94.02	97.59%	97.07–98.11	98.14%	97.69–98.59
Specificity	98.07%	97.61–98.53	97.96%	97.48–98.44	94.69%	93.94–95.44	98.98%	98.64–99.32	98.67%	98.28–99.06	98.45%	98.03–98.87
F1-score	92.10%	91.19–93.01	94.39%	93.62–95.16	96.55%	95.94–97.16	95.03%	94.30–95.76	97.70%	97.20–98.20	98.55%	98.15–98.95
Accuracy	96.52%	95.90–97.14	95.85%	95.18–96.52	95.85%	95.18–96.52	97.82%	97.33–98.31	98.26%	97.82–98.70	98.26%	97.82–98.70
AUC	98.94%	98.60–99.28	99.20%	98.90–99.50	99.15%	98.84–99.46	99.27%	98.98–99.56	99.80%	99.65–99.95	99.83%	99.69–99.97
Training time	6 h. 30 min. 42 s		6 h. 12 min. 06 s		6 h. 23 min. 47 s		10 h. 54 min. 39 s		10 h. 41 min. 59 s		11 h. 52 min. 48 s	

T2-weighted images

Statistic	ResNet-50						EfficientNet-b0					
	CNO		OM		TR		CNO		OM		TR	
	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI
Sensitivity	94.88%	94.14–95.62	83.60%	82.35–84.85	97.96%	97.48–98.44	95.63%	94.94–96.32	85.22%	84.03–86.41	98.56%	98.16–98.96
Specificity	98.56%	98.16–98.96	97.47%	96.94–98.00	96.74%	96.14–97.34	99.06%	98.74–99.38	98.46%	98.05–98.87	98.72%	98.34–99.10
F1-score	94.93%	94.19–95.67	89.07%	88.02–90.12	97.92%	97.44–98.40	96.17%	95.52–96.82	90.80%	89.83–91.77	98.86%	98.50–99.22
Accuracy	97.73%	97.23–98.23	92.18%	91.28–93.08	97.48%	96.95–98.01	98.29%	97.85–98.73	93.42%	92.59–94.25	98.63%	98.24–99.02
AUC	99.08%	98.76–99.40	97.08%	96.51–97.65	99.64%	99.44–99.84	99.32%	99.04–99.60	97.87%	97.38–98.36	99.91%	99.81–99.96
Training time	6 h. 22 min. 04 s		6 h. 21 min. 29 s		6 h. 21 min. 21 s		11 h. 15 min. 01 s		11 h. 49 min. 04 s		11 h. 19 min. 55 s	

97.47%, and 96.74%, respectively. EfficientNet-b0 exhibits a sensitivity of 93.17% for CNO in T1-weighted images and 95.63% in T2-weighted images. It also achieves sensitivity percentages of 97.59% and 85.22% for OM and 98.14% and 98.56% for TR in T1- and T2-weighted images, respectively. Regarding specificity, CNO, OM, and TR values in T1-weighted cases are 98.98%, 98.67%, and 98.45%, respectively. Similarly, for T2-weighted cases, EfficientNet-b0 yields specificity of 99.06%, 98.46%, and 98.72% for CNO, OM, and TR, respectively. Performance metrics and training times are provided in Table 2 for T1- and T2-weighted images. The formulas for the metrics are provided in detail in the Supplementary File.

Discussion

According to the MCC classification results, both ResNet-50 and EfficientNet-b0 exhibit high accuracies of 96.17%, 95.57%, 97.11%, and 96.83% for T1- and T2-weighted images, respectively. When assessing the performance of deep learning–based algorithms using both BCC and MCC metrics (as shown in Table 1 and 2), it is evident that each strategy achieves remarkable accuracy in classification. This underscores the capability of both ResNet-50 and EfficientNet-b0 to effectively discern between CNO, OM, and TR classes utilizing T1- and T2-weighted MR images.

Approximately 40–60 million people worldwide with diabetes suffer from diabetic foot complications, 50% have a risk of getting OM, and the prevalence of CNO in people with diabetes with neuropathy has increased to up to 35% [30]. Diabetic foot complications can encompass a range of conditions, such as CNO, OM, and soft tissue complications, which may include cellulitis, myositis, ulceration, sinus tracts, and abscesses. Distinguishing between BMSAs associated with OM and CNO is essential, particularly in diabetic foot complications. OM is a severe bacterial infection of the bone, requiring prompt and targeted antibiotic treatment to prevent severe consequences, including the risk of limb amputation. CNO is a non-infectious condition characterized by bone and joint deformities resulting from nerve damage, often associated with diabetes. Misdiagnosing OM as Charcot neuropathy or vice versa can lead to inappropriate treatment choices, potentially exacerbating the condition. In diabetic individuals who are already at an increased risk of foot complications, accurate differentiation is crucial for guiding appropriate interventions. Mismanagement can result in delayed or inadequate treatment, contributing to prolonged healing times, increased morbidity, and elevated healthcare costs. Morphologic MR imaging is considered the most valuable diagnostic method for assessing diabetic foot complications [31]. Despite its usefulness, differentiating between Charcot neuroarthropathy and osteomyelitis using MR imaging can be challenging [3, 4]. Moreover, the treatment strategies for diabetic foot complications differ significantly, specifically infection and Charcot neuroarthropathy. The primary treatment approach involves a combination of antibiotics and surgical debridement for infections such as OM or soft tissue infections.

On the other hand, the mainstay of treatment for CNO is protected weight-bearing [1]. Also, the location of the signal abnormality and soft tissue findings on imaging can be crucial diagnostic features. Osteomyelitis typically occurs due to the contiguous spread of infection from an adjacent skin ulceration or soft tissue infection to the underlying bone [32]. CNO typically involves multiple midfoot bones, and the affected bones may show marrow abnormalities in the periarticular and subchondral areas [33].

In recent years, deep learning techniques have become increasingly important in the prognosis and diagnosis of diseases across a wide range of medical imaging modalities. Although numerous studies were published in the field of DFU, to our knowledge, only one study focused on classifying CNO, OM, and TR on MRI images utilizing radiomics features [26]. Cuce et al. first extracted radiomic features from the original, Laplacian of Gaussian (LoG) filtered, and the wavelet decomposed versions of T1- and T2-weighted MRI images and then applied a two-layer cascade feature selection method to determine the optimal feature set. The multi-layer perceptron (MLP) achieved the best classification results, with an accuracy of 76.92% for T1-weighted images and 84.38% for T2-weighted images. Although the radiomic-based algorithm achieved a reasonable accuracy for T2-weighted images, the proposed DL-based classifiers have better and similar performance for both T1- and T2-weighted images.

From Table 1, for T1-weighted images, 61 of the CNO cases were misclassified as OM and TR for ResNet-50, and 27 of the CNO cases were misclassified as EfficientNet-b0. Also, ResNet-50 misclassified 19 of OM, and EfficientNet-b0 misclassified 40 of OM as CNO and TR, respectively. For TR, the number of misclassified cases is 50 and 31 for ResNet-50 and EfficientNet-b0, respectively. In the case of T2-weighted images, ResNet-50 misclassified 79 samples of CNO as OM and TR, while EfficientNet-b0 misclassified 45 instances of CNO. Similarly, ResNet-50 misclassified 44 samples of OM, and EfficientNet-b0 incorrectly labeled 33 as CNO and TR samples, respectively. For TR cases, there were 35 misclassifications for ResNet-50 and 35 for EfficientNet-b0. The classification performance of the CNO, OM, and TR cases has approximately the same AUC values (from Table 2 T1-weighted images; minimum AUC was 98.94%, and maximum AUC was 99.83%). For T2-weighted MRI, the classification performance of the CNO and the TR cases has nearly the same and better AUC values than in OM cases (from Table 2 T2-weighted images;

compared to the CNO and TR AUC values, the OM AUC values were 2% lower).

As a limitation of our study, it is important to note that the dataset size was constrained, and we recognize the potential correlation among slices from the same patient. Despite this recognition, we adopted a practical approach based on the determination of experienced radiologists, treating each slice as an independent observation within the context of our analysis. The inherent correlation among slices, arising from the fact that samples from the same patient belong to the same type of bone marrow signal abnormality (belonging to the same class), introduces a nuanced consideration. Secondly, the utilization of semi-automatic segmentation in our study may be a limitation. Because in the deep learning literature, DL-based object detection algorithms [34–36] are commonly employed, involving both localization and classification components to identify ROI and categorize objects. In this study, segmentation of BMSA regions could have been performed fully automatically using object detection algorithms instead of a semi-automatic method. However, false-positive regions and relatively lower segmentation accuracy may impact the overall success of the classification process. Because classification in object detection algorithms depends on segmentation, false localizations are the primary source of errors [37].

In this study, we assessed the capability of deep learning algorithms for differentiating the signal intensity of bone marrow among CNO, OM, and TR cases, considering T1 and T2 images of the cases. Consequently, this study represented the first attempt to differentiate bone marrow signal abnormalities using deep learning methods. When the classification performance of the algorithms was considered, EfficientNet-b0 and ResNet-50 demonstrated that DL could serve as a non-invasive tool to differentiate CNO, OM, and TR with the phenomenon of transfer learning for T1- and T2-weighted MRI images.

Declarations

Conflict of Interest The authors declare no competing interests.

References

1. J. C. Baker, J. L. Demertzis, N. G. Rhodes, D. E. Wessell and D. A. Rubin, "Diabetic musculoskeletal complications and their imaging mimics," *Radiographics,* vol. 32, no. 7, pp. 1959-74, 2012.

2. P. D. Brash, J. Foster, W. Vennart, P. Anthony, J. E. Tooke, "Magnetic resonance imaging techniques demonstrate soft tissue damage in the diabetic foot," *Diabetic Medicine,* 16(1), 55-61, 1999.

3. F. B. Ergen, S. E. Sanverdi and A. Oznur, "Charcot foot in diabetes and an update on imaging," *Diabet Foot Ankle,* vol. 20, no. 4, pp. 124-127, 2013.

4. L. C. Rogers and N. J. Bevilacqua, "Imaging of the Charcot foot," *Clinics in Podiatric Medicine and Surgery,* vol. 25, p. 263–74, 2008.

5. T. Martín Noguerol, A. Luna Alcalá, L. S. Beltrán, M. Gómez Cabrera, J. Broncano Cabrero and J. C. Vilanova, "Advanced MR Imaging Techniques for Differentiation of Neuropathic Arthropathy and Osteomyelitis in the Diabetic Foot," *Radiographics,* vol. 37, no. 4, pp. 1161–1180, 2017.

6. S. Iqbal, A. N. Qureshi, J. Li and T. Mahmood, "On the Analyses of Medical Images Using Traditional Machine Learning Techniques and Convolutional Neural Networks," *Archives of Computational Methods in Engineering,* 2023.

7. A. Han, Y. Zhang, A. Li, C. Li, F. Zhao, Q. Dong, Y. Liu, X. Shen, S. Yan, S. Zhou, "Deep Learning Methods for Real-time Detection and Analysis of Wagner Ulcer Classification System," *2022 International Conference on Computer Applications Technology (CCAT) IEEE*, pp. 11–21, 2022.

8. K. Bousabarah, M. Ruge, J. S. Brand, M. Hoevels, D. Rueß, J. Borggrefe, N. G. Hokamp, V. Visser-Vandewalle, D. Maintz, H. Treuer and M. Kocher, "Deep convolutional neural networks for automated segmentation of brain metastases trained on clinical data," *Radiation Oncology,* 2020.

9. M. Goyal, N. Reeves, A. Davison, S. Rajbhandari, J. Spragg and M. Yap, "Dfunet: Convolutional neural networks for diabetic foot ulcer classification," *arXiv,* 2017.

10. I. Cruz-Vega, D. Hernandez-Contreras, H. Peregrina-Barreto, J. Rangel-Magdaleno and J. Ramirez-Cortes, "Deep learning classification for diabetic foot thermograms," *Sensors,* 2020.

11. K. Munadi, K. Saddami, M. Oktiana, R. Roslidar, K. Muchtar, M. Melinda, R. Muharar, T. Abidin and F. Arnia, "A Deep Learning Method for Early Detection of Diabetic Foot Using Decision Fusion and Thermal Images," *Applied Sciences,* 2022.

12. A. I. G. Diez, D. Fuster, L. Morata, F. Torres, R. Garcia, D. Poggio, S. Sotes, M. Del Amo, J. Isern-Kebschull, J. Pomes, A. Soriano, L. Brugnara and X. Tomas, "Comparison of the diagnostic accuracy of diffusion-weighted and dynamic contrast-enhanced MRI with 18F-FDG PET/CT to differentiate osteomyelitis from Charcot neuro-osteoarthropathy in diabetic foot," *European Journal of Radiology,* 2020.

13. M. Goyal, N. D. Reeves, S. Rajbhandari, N. Ahmad, C. Wang and M. H. Yap, "Recognition of ischaemia and infection in diabetic foot ulcers: Dataset and techniques," *Computers in Biology and Medicine,* vol. 117, 2020.

14. M. H. Yap, R. Hachiuma, A. Alavi, C. R. B. Brüngel, M. Goyal, H. Zhu, J. Rückert, M. Olshansky, X. Huang, H. Saito, S. Hassanpour, C. M. Friedrich, D. B. Ascher, A. Song, H. Kajita and D. Gill, "Deep learning in diabetic foot ulcers detection: A comprehensive evaluation," *Computers in Biology and Medicine,* vol. 135, 2021.

15. B. Cassidy, N. D. Reeves, J. M. Pappachan, D. Gillespie, C. O'Shea, S. Rajbhandari, A. G. Maiya, E. Frank, A. J. Boulton, D. G. Armstrong, B. Najafi, J. Wu, R. S. Kochhar and M. H. Yap, "The DFUC 2020 Dataset: Analysis Towards Diabetic Foot Ulcer Detection," *TouchREVIEWS in endocrinology,* vol. 17, no. 1, pp. 5-11, 2021.

16. M. Goyal and S. Hassanpour, "A Refined Deep Learning Architecture for Diabetic Foot Ulcers Detection," *Computer Science,* 2020.

17. A. Hernandez-Guedes, I. Santana-Perez, N. Arteaga-Marrero, H. Fabelo, G. M. Callico and J. Ruiz-Alzola, "Performance

Evaluation of Deep Learning Models for Image Classification Over Small Datasets: Diabetic Foot Case Study," *IEEE Access,* vol. 10, pp. 124373-124386, 2022.

18. S. Muralidhara, A. Lucieri, A. Dengel and S. Ahmed, "Holistic multi-class classifcation & grading of diabetic foot ulcerations from plantar thermal images using deep learning," *Health Information Science and Systems,* vol. 10, no. 21, 2022.

19. M. Ahsan, S. Naz, R. Ahmad, H. Ehsan and A. Sikandar, "A Deep Learning Approach for Diabetic Foot Ulcer Classification and Recognition," *Information,* vol. 14, no. 1, p. 36, 2023.

20. L. Alzubaidi, M. A. Fadhel, S. R. Oleiwi, O. Al-Shamma and J. Zhang, "DFU_QUTNet: diabetic foot ulcer classification using novel deep convolutional neural network," *Multimedia Tools and Applications,* vol. 79, pp. 15655-15677, 2019.

21. P. N. Thotad, G. R. Bharamagoudar and B. S. Anami, "Diabetic foot ulcer detection using deep learning approaches," *Sensors International,* vol. 4, 2023.

22. A. Anaya-Isaza and M. Zequera-Diaz, "Detection of Diabetes Mellitus With Deep Learning and Data Augmentation Techniques on Foot Thermography," *IEEE Access,* vol. 10, pp. 59564-59591, 2022.

23. T. K. Chuah, E. Van Reeth, K. Sheah, C. L. Poh, "Texture analysis of bone marrow in knee MRI for classification of subjects with bone marrow lesion—data from the Osteoarthritis Initiative," *Magnetic Resonance Imaging,* 31(6), 930-938, 2013.

24. J. Li, S. Fu, Z. Gong, Z. Zhu, D. Zeng, P. Cao, T. Lin, T. Chen, X. Wang, R. Lartey, C. K. Kwoh, A. Guermazi, F. W. Roemer, D. J. Hunter, J. Ma, C. Ding, "MRI-based texture analysis of infrapatellar fat pad to predict knee osteoarthritis incidence, " *Radiology,* 304(3), 611-621, 2022.

25. S. Kostopoulos, N. Boci, D. Cavouras, A. Tsagkalis, M. Papaioannou, A. Tsikrika, D. Glotsos, P. Asvestas, E. Lavdas, "Radiomics Texture Analysis of Bone Marrow Alterations in MRI Knee Examinations," *Journal of Imaging*, 9(11), 252, 2023.

26. F. Cuce, G. Tulum, K. B. Yilmaz, O. Osman and A. Aralasmak, "Radiomics method in the differential diagnosis of diabetic foot osteomyelitis and charcot neuroarthropathy," *The British Journal of Radiology,* 2023.

27. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 770–778, 2016.

28. M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *International Conference on Machine Learning,* 2019.

29. G. Tulum, "GitHub," [Online]. Available: https://github.com/DrGokalpTulum/deep_learning_classification_CNO_OM_TR/tree/main. [Accessed 16 8 2023].

30. B. A. Lipsky, É. Senneville, Z. G. Abbas, J. Aragón-Sánchez, M. Diggle, J. M. Embil, S. Kono, L. A. Lavery, M. Malone, S. A. Van Asten, V. Urbančič-Rovan, E. J. G. Peters, "Guidelines on the diagnosis and treatment of foot infection in persons with diabetes (IWGDF 2019 update)," *Diabetes/metabolism research and reviews,* 2020.

31. K. T. Low, W. C. Peh, "Magnetic resonance imaging of diabetic foot complications, " *Singapore medical journal,* 56(1), 23, 2015.

32. H. P. Ledermann, W. B. Morrison and M. E. Schweitzer, "MR image analysis of pedal osteomyelitis: distribution, patterns of spread, and frequency of associated ulceration and septic arthritis," *Radiology,* vol. 223, no. 3, pp. 747-755, 2002.

33. C. Yang and A. Tandon, "A Pictorial Review of Diabetic foot Manifestations," *The Medical journal of Malaysia,* vol. 68, no. 3, p. 279–289, 2013.

34. R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *2014 IEEE Conference on Computer Vision and Pattern Recognition,* 2013.

35. R. Girshick, "Fast R-CNN," *2015 IEEE International Conference on Computer Vision,* pp. 1440–1448, 2015.

36. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," *2016 IEEE conference on computer vision and pattern recognition,* pp. 779–788, 2016.

37. R. Sujatha, S. L. Aarthy and R. R. Vettriselvan, "Integrating Deep Learning Algorithms to Overcome Challenges in Big Data Analytics," *CRC Press,* 2021.