Recep Tayyip Erdoğan Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 4(2) 288-305 2023
Recep Tayyip Erdogan University Journal of Science and Engineering, 4(2) 288-305 2023
ISSN: 2687-2315 / E-ISSN: 2757-7686

Differential Diagnosis of Diabetic Foot with Deep Learning Methods

Maide ÇAKIR[1*], Hüseyin CANBOLAT[2], Gökalp TULUM[3]

[1] Ankara Yıldırım Beyazıt University, Graduate School of Natural and Applied Sciences, Department of Electrical and Electronic Engineering, Türkiye

[2] Ankara Yıldırım Beyazıt University, Faculty of Engineering and Natural Sciences, Department of Electrical and Electronic Engineering, Türkiye

[3] İstanbul Topkapı University, Faculty of Engineering, Department of Electrical and Electronic Engineering, Türkiye

[*]Sorumlu Yazar/Corresponding Author
E-mail: maidecakrr@gmail.com

Derleme Makalesi/Review Article
Geliş Tarihi/Received:17.10.2023
Kabul Tarihi/Accepted: 02.11.2023

ABSTRACT

Diabetic foot complications, caused by prolonged hyperglycemia, are a significant health concern among diabetes patients. The majority of patients develop diabetic foot complications, contributing significantly to diabetes-related hospital admissions. These complications include foot ulcers, infections, ischemia, Charcot foot, and neuropathy. They also increase the risk of amputation, affecting quality of life and putting strain on healthcare systems. At this stage, early diagnosis plays a vital role. The process of diagnosing involves not only identifying the presence or absence of a disease, but also categorizing the disease. In this study, we examine the use of deep learning methods in the diagnosis of diabetic foot conditions. It explores various aspects, such as predictive modeling and image analysis. The study discusses the progression of model designs, data sources, and interpretability methodologies, with a focus on improving accuracy and early detection. Overall, the study provides a comprehensive analysis of the current state of deep learning in diabetic foot problems, with highlighting advancements.

Keywords: Deep learning, differential diagnosis, diabetic foot, classification.

Diyabetik Ayağın Derin Öğrenme Yöntemleriyle Ayırıcı Tanısı

ÖZ

Uzun süreli hipergliseminin neden olduğu diyabetik ayak komplikasyonları diyabet hastaları arasında önemli bir sağlık sorunudur. Hastaların çoğunda diyabetik ayak komplikasyonları gelişir ve bu da diyabetle ilişkili hastaneye başvurulara önemli ölçüde sebebiyet verir. Bu komplikasyonlar arasında ayak ülserleri, enfeksiyonlar, iskemi, Charcot ayağı ve nöropati yer alır. Ayrıca amputasyon riskini artırarak yaşam kalitesini etkiler ve sağlık sistemleri üzerinde baskı yaratır. Bu aşamada erken teşhis hayati önem taşır. Teşhis süreci yalnızca bir hastalığın varlığını veya yokluğunu belirlemeyi değil aynı zamanda hastalığın kategorize edilmesini de içerir. Bu çalışmada diyabetik ayak rahatsızlıklarının tanısında derin öğrenme yöntemlerinin kullanımı incelenmiştir. Çalışma, tahmine dayalı modelleme ve resim analizi de dahil olmak üzere farklı yönleri de ele alır. Doğruluğun ve erken tespitin geliştirilmesine odaklanarak model tasarımlarının, veri kaynaklarının ve yorumlanabilirlik metodolojilerinin ilerleyişini tartışır. Genel olarak bu çalışma, diyabetik ayak problemlerinde derin öğrenmenin mevcut durumunun kapsamlı bir analizini ve ilerlemelerin altını çizmektedir.

Anahtar Kelimeler: Derin öğrenme, ayırıcı tanı, diyabetik ayak, sınıflandırma.

Cite as;
Çakır, M., Canbolat, H., Tulum, G. (2023). Differential diagnosis of diabetic foot with deep learning methods, *Recep Tayyip Erdogan University Journal of Science and Engineering,* 4(2), 288-305. DOI: 10.53501/rteufemud.1377390

1. Introduction

Diabetes mellitus (DM) is a chronic condition characterized by compromised insulin production, insulin resistance, or a combination of both (Belsti et al., 2020). According to the International Diabetes Federation, there is an estimated global prevalence of diabetes of approximately 537 million individuals (Maltese et al., 2023). According to the cited source, it is projected that the aforementioned figure will increase to approximately 783 million by the year 2045 (Ogurtsova et al., 2022). Nevertheless, it is worth noting that around 50% of diabetes cases remain misdiagnosed, leading to a staggering 6.7 million fatalities. Undiagnosed individuals with diabetes are at a significantly increased risk of developing a range of serious diseases (Ferreira et al., 2020). DM is accompanied by several consequences, such as cardiovascular diseases, myocardial infarctions, renal diseases, retinopathy, podiatric injuries, hearing impairment, visual impairments, bacterial and fungal infections, depressive disorders, and dementia. Figure 1 shows the principal problems associated with diabetes. These challenges not only have implications for individuals' well-being, but they also have a substantial influence on their personal and professional lives. Additionally, it is linked to a significant mortality rate (Cruz-Vega et al., 2020).

Diabetic foot is a prominent problem related to diabetes. According to reports, individuals with diabetes may have reduced sensitivity in their feet, along with mechanical stress in the plantar area. The occurrence of diabetic foot is a significant consequence associated with diabetes, characterized by the formation of plantar ulcers that may ultimately need amputation (Cruz-Vega et al., 2020). Around one-third of individuals diagnosed with diabetes will experience a diabetic foot ulcer (DFU), with a lifetime probability of 34%. This indicates a significant health risk for patients (Yap et al., 2021b). Individuals diagnosed with DFU are susceptible to experiencing compromised wound healing. DFUs have been associated with the potential risk of lower limb amputation and a subsequent decrease in survival rates (Chamberlain et al., 2022). Furthermore, it is crucial to note that the primary risk factors associated with the formation of foot ulcers in individuals diagnosed with diabetes are peripheral neuropathy and vascular disease (Reardon et al., 2020).

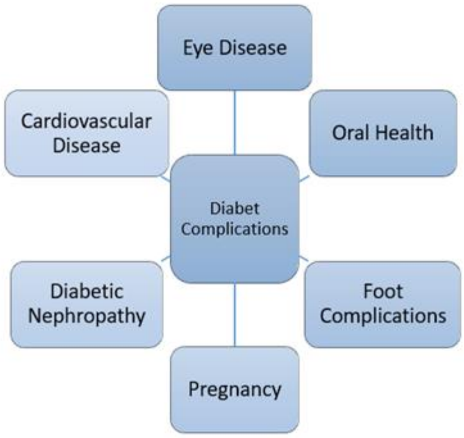

Figure 1. Diabetes related disorders (Das et al., 2023)

Diabetes, especially DFU management, is costly due to the expenses associated with diagnosis, regular check-ups, costly substances, and maintaining personal hygiene, as shown in Figure 2. In recent years, there has been widespread adoption of medical technology with a view of enhancing the care provided to individuals with diabetes. This has resulted in the generation of a substantial volume of data, which holds potential for advancing the management of such chronic condition. Considering this opportunity, methodologies that employ artificial intelligence (AI), have been extensively embraced with encouraging outcomes (Zhu et al., 2020). Automated telemedicine systems have become the most economical option for detecting DFUs. Identification methods, such as computer vision techniques and supervised machine learning (ML) and deep learning (DL) algorithms, have been suggested for this purpose (Das et al., 2023). The use of AI approaches to various medical imaging has become increasingly prevalent with the advancement of AI technology. ML and DL, as widely adopted methods in the field of AI, have established their dominance over an extended period.

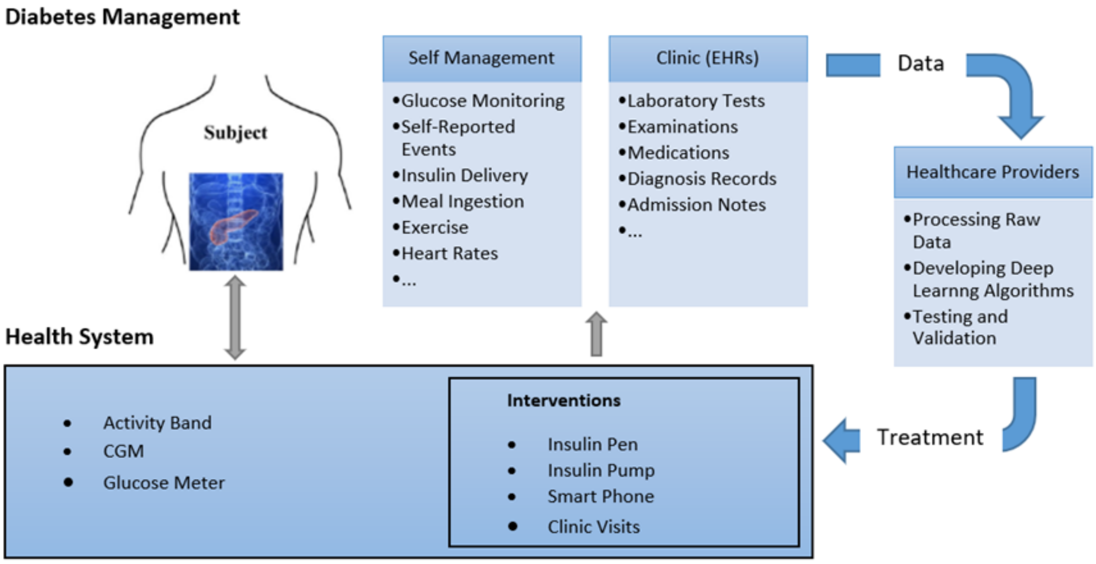

Figure 2. Diabetes management (Zhu et al., 2020)

The fundamental principle behind several DL methods is the use of artificial neural networks (ANNs) with extensive architectures to process input and acquire progressively more complex properties. It is currently used to improve classification accuracy, but it often requires a large amount of data and labeling. Two primary methods for training DL are full training, which requires a large amount of labeled data, and using large computing and memory resources like GPUs for accelerated processing. In medicine, this may pose challenges for expert annotation tasks and acquiring large patient images. One potential solution is fine-tuning a pre-existing DL architecture using a large, labeled dataset from a specific application domain. In medical imaging, fine-tuning has shown superior performance or comparable performance to a convolutional neural network (CNN) trained from the beginning (Cruz-Vega et al., 2020).

2. Material and Method

In recent years, there has been a significant advancement in the field of DL, which is a specific branch of ML. In contrast to traditional ML approaches that necessitate human feature extraction and rely on domain expertise, DL could automatically extract features by shifting from hand-designed to data-driven features (Min et al., 2017). Figure 3 illustrates the distinction between traditional ML and DL in regards to feature extraction. In conventional ML, the process of feature extraction often involves multiple stages, including feature extraction and selection. DL is a computational framework that typically consists of numerous layers of processing. Its purpose is to acquire knowledge about data by converting input information into various degrees of abstraction. This is achieved through the use of basic yet non-linear modules (LeCun et al., 2015). Through these transitions, DL models will acquire knowledge of a highly intricate function. Significantly, because of its automated nature, DL facilitates the analysis of numerous cases, surpassing the capacity of human experts in terms of both exposure and recollection. Consequently, DL exhibits enhanced resilience to diverse variances in characteristics across distinct categories (Chan et al., 2020).

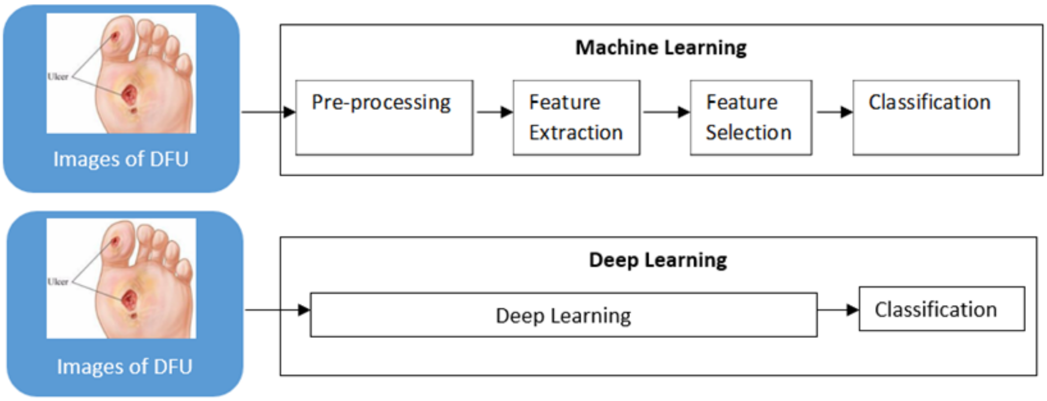

Figure 3. Difference between DL and ML (Zhang et al., 2022)

DL models provide a flexible and adjustable output with high accuracy, minimizing costs, reducing human bias, and reducing time-consuming work. However, their successful application in healthcare settings has not been fully realized. The data challenge, or insufficiently labeled data, is a primary obstacle in this field. DL models often exhibit a large parameter count and overfitting tendencies when trained on insufficiently large datasets, leading to suboptimal performance on novel data. Various strategies, such as transfer learning and data augmentation, can address this issue. The prevailing trend towards deep neural networks (DNN) intensifies the vanishing gradient problem. Skip connections have been demonstrated to address this issue and offer additional advantages during the training phase.

DL algorithms could be categorized into reinforcement learning, supervised learning, and unsupervised learning. Supervised learning involves classification and regression tasks, using labeled input data for backward propagation and model optimization. Three types of supervised learning based DNNs have been identified in the diabetes literature: CNNs, deep multilayer perceptrons (DMLPs), and recurrent neural networks (RNNs). DMLPs, also known as feed-forward neural networks, utilize simple interconnections among neurons, specifically fully connected (FC) layers. DMLPs are characterized by weight vectors, bias scalars, and nonlinear activation functions. CNNs use convolutional layers as perceptrons to analyze signals from multi-dimensional arrays, leading to exceptional performance in imaging tasks. Most CNN architectures use a subsampling layer or pooling layer to aggregate feature maps. Convolutional processes reduce neuron connections across layers, improving model training efficiency by facilitating back-propagation. Various CNN-based methods have been used for large-scale imaging recognition tasks, such as the ImageNet database, using parallelized operations of GPUs and TPUs. These models have also been adapted for practical use in industry. Common CNN configurations include VGGNet, AlexNet, ResNet, EfficientNet, and GoogLeNet (Zhu et al., 2020).

2.1. CNN

CNNs are crucial in medical diagnostic applications, specifically for analyzing and interpreting medical imaging data like magnetic resonance imaging (MRIs), X-rays, computerized tomography (CT) scans, and histopathological images. A conventional CNN structure consists of convolutional layers, dropout layers, pooling layers, and FC layers. Convolutional layers extract feature maps from input images by multiplication with a convolution kernel matrix (Roback, 2010). Non-linear characteristics are derived from convolved outputs using non-linear activation functions (Yap et al., 2021a). Pooling layers decrease feature map resolution and obtain invariance. To mitigate overfitting and minimize computational complexity, techniques like max-min, and average pooling are used. The feature matrices in the last pooling layer are flattened and converted to a vector for the FC layer. The classification label is determined by applying a transformation to the flattened vector. The softmax function generates a probability distribution for multiple classifications. Local patterns are found by convolutional layers, spatial dimensions are reduced by pooling layers, and features are combined by FC layers, and the final diagnosis is made by the output layer. The network is trained using annotated datasets and optimized using methods like stochastic gradient descent. This architecture aids healthcare practitioners in making accurate, prompt, and potentially life-saving evaluations (Zhang et al., 2022). Figure 4 shows the CNN architecture based on the image classification of DFU.

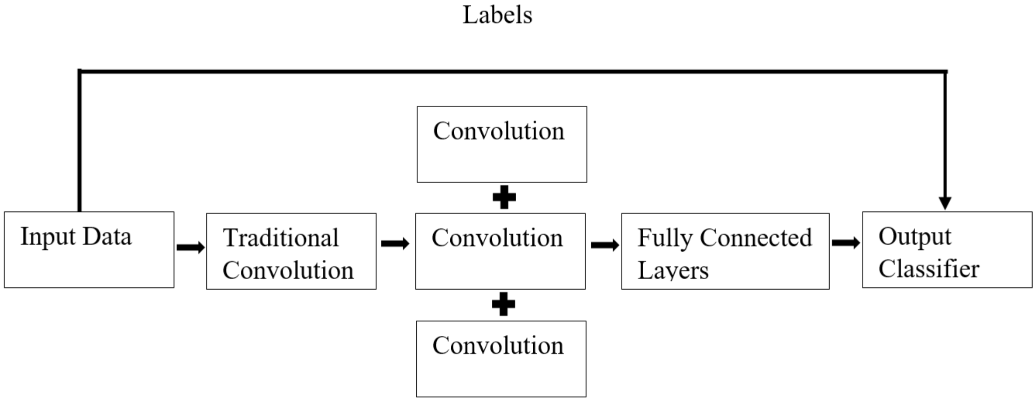

Figure 4. CNN architecture for image classification of DFU (Zhang et al., 2022)

Amin et al. (2020) present a DL model for localizing and classifying DFU using a 16-layer CNN architecture. The model includes various components such as 1 input layer, 3 convolutional layers, 3 batch-normalization layers, 1 average pooling layer, 1 skip convolutional layer, 3 ReLU layers, 1 add layer (performing element-wise addition of two inputs), FC layers, SoftMax layer, and classification output layers. It incorporates YOLOv2-DFU for infection/ischemia models. Various classifiers, including K-Nearest Neighbors, Decision Trees, Ensemble methods, SoftMax, and Naive Bayes, are used to analyze the classification results. The YOLOv2-DFU model incorporates a shuffle network for localization, demonstrating superior performance due to the influence of the chosen convolutional layers on classification outcomes. Das et al. (2022) present a novel approach for identifying DFU using a stacked parallel convolution layer within a CNN. The DFU SPNet, trained with various optimizer and learning rate configurations, demonstrated superior performance compared to standard CNN architectures. The DFU SPNet demonstrated significant gains over state-of-the-art approaches on the same dataset, with promising mean average accuracy, sensitivity, F1-score, and AUC. This approach, DFU SPNet, can assist DFU specialists in expediting their decision-making process. Harahap et al. (2022), evaluated the CNN algorithm, ResNet152V2, for its precision in categorizing diabetic ulcer disease using transfer learning methodology. The ResNet152V2 model achieved the highest accuracy score of 0.993, recall score of 0.986, precision score of 1.00, F1-Score of 0.993, and support score of 72. This study highlightsm CNN's potential as a classifier for identifying and categorizing diabetic ulcers in individuals with DM. Gamage et al. (2019) propose the use of a CNN based on the DenseNet-201 architecture for predicting the severity stages of DFUs. The CNN is combined with a global average pooling (GAP) layer to extract features. The authors also introduce a methodology that improves computational efficiency and reduces memory usage by implementing feature extraction using singular value decomposition. The proposed architecture has the potential to achieve an accuracy rate of over 96%. The researchers evaluate various pre-trained CNN architectures, including ResNet, EfficientNet, DenseNet, Xception, VGG, InceptionV3, and InceptionResNetV2, commonly used for transfer learning. Reyes-Luévano et al. (2023) present a unique architecture called DFU_VIRNet, which is a CNN designed for the purpose of automatically classifying abnormal skin from healthy skin. The proposed CNN is trained and validated using images obtained from thermography, specifically in the visible and infrared spectrums. In addition, this study introduces a novel approach utilizing estimation maps for the identification of high-risk areas where patients are more likely to develop DFU. In order to assess the applicability of the proposed method, the performance of DFU_VIRNet has been evaluated with a total of five datasets. These datasets consisted of one visible-infrared dataset and four recently published visible datasets, as presented in this paper. The datasets encompassed samples of abnormal skin, healthy skin, as well as cases of ischaemia and infection. The findings of the study indicate that DFU_VIRNet exhibited superior performance compared to the current leading results. Specifically, it achieved an AUC of 0.9923 and an F1-score of 0.9600 for DFU classification, an AUC of 0.9982 and an F1-score of 0.9928 for ischaemia classification, and an AUC of 0.9121 and an F1-score of 0.8363 for infection classification. The exceptional performance of DFU_VIRNet can be attributed to a novel learning mechanism introduced in this study, known as the GAP-2D-DLSA-IMG substructure. This mechanism effectively mitigates overfitting and enhances the perceptual field of DFU_VIRNet.

2.1.1. R–CNN

The use of R-CNN (Region-based CNN) in diabetic foot diagnosis has shown significant progress in object detection within medical images. R-CNN models are known for their high accuracy in localizing and detecting objects, particularly in identifying anomalies related to diabetic foot conditions such as ulcers and infections. These models employ a two-stage process of region proposal generation and CNN-based classification, improving accuracy for evaluation and treatment planning in medical imaging. Customization through fine-tuning on specialized datasets specific to diabetic foot further enhances diagnostic performance. The presentation of interpretable outcomes, represented by bounding boxes around identified areas, allows healthcare practitioners to visually verify and understand model predictions. Incorporating R-CNN models into diagnostic systems improves the efficiency of automated analysis for diabetic foot images, leading to timely treatments and reducing the severity of foot issues. These models also complement conventional DL methods in classification and object identification, making them crucial components of comprehensive diagnostic solutions. However, it is important to note that the effectiveness of R-CNN in diagnosing diabetic foot relies on carefully annotated medical image datasets and rigorous training and validation procedures.

In contrast to conventional feed-forward neural networks, RNNs incorporate information from past timesteps into their input. The inclusion of this feature enhances the capabilities of RNNs in effectively analyzing consecutive data and capturing temporal characteristics. Nevertheless, the challenge with vanilla RNNs resides in the process of back-propagation training, which often leads to issues such as exploding and gradient vanishing (Bengio et al., 1994). Long short-term memory and gated recurrent units are two examples of sophisticated RNN cells that have fortunately addressed the issues mentioned above. These RNN cells have successfully addressed the issues by incorporating gate functions and effectively retaining long-term information. RNN-based models have demonstrated significant advancements in many prediction and regression tasks, particularly in the domains of natural language processing and speech recognition. The attention mechanism has emerged as a prominent trend in RNNs. This mechanism enables models to selectively concentrate on specific segments of input sequences, facilitating the mapping of dependencies irrespective of their spatial separation (Zhu et al., 2020).

Oliveira et al. (2021) present a novel methodology for the automated detection of DFUs via DL methodologies. An expanded iteration of the Faster R-CNN methodology has been implemented in the study. Several strategies have been implemented in order to attain a high level of precision in the detection of ulcers, reduce the occurrence of false positive results, and accelerate the whole detection process. Several modifications were made to improve the performance of the detection system. These modifications include altering the number of regions and anchor scales, implementing data augmentation techniques in the dataset, and adopting a CNN that has demonstrated superior detection capabilities compared to previous approaches. Ultimately, they conducted experiments with the selected detectors, subjecting each one to a training process consisting of 100 epochs. The findings of our study indicate that the application of our solutions leads to enhancements in both the mean average precision (mAP) and F1-score, in comparison to the state-of-the-art standard detector implementations. Significant improvements have been made in terms of mAP, F1-score, and detection speeds. These advancements not only enhance the detection capabilities of DFUs, but also instill greater confidence in the practical implementation of the Faster R-CNN DFU model. Cao et al. (2023) used the Wagner diabetic foot grading method to categorize wounds into five grades using a multi-classification methodology. An ADL model was developed to perform semantic segmentation of DFU wounds, based on the Mask Region-based CNN (Mask R-CNN) architecture. The model achieved several levels of diabetes nested segmentation outcomes, effectively capturing varying degrees of severity within a single wound. The model demonstrated superior performance metrics, with a specificity of 99.50%, sensitivity of 70.62%, precision of 84.56%, and a mAP of 85.70%. This approach offers recommendations for the assessment, diagnosis, and management of DFUs, demonstrating the efficacy of the nested segmentation and multi-level classification approaches. Sharma et al. (2023) explore the use of AI and image fusion techniques to assess and characterize DFUs. The study utilizes a computer-aided assessment technique that combines thermal and visual data using an image fusion technique based on hue, saturation, and value. The wound area estimation is achieved through a Mask R-CNN based on instance segmentation. The study is a randomized, prospective, single-blind trial conducted over a period of 12 weeks, focusing on neuropathic DFUs of Wagner grade 2 located on the plantar surface of the foot. Forty-two participants with an average age of 54.28±7.45 years and an average ulcer duration of 5.86±2.22 years were included in the study. The healing progress of eight patients was tracked on a weekly basis. The study found that the absolute temperature differential across contralateral ulcer sites was measured to be 2.63±1.99°C, with a statistically significant p-value of 0.000040412. There was a 92.50% correlation between the ground truth ulcer area estimation made by doctors using the Woundly program and the suggested method. The research concludes that the Mask-RCNN technique, when applied to fused images, has the potential to enable automation and user-independence.

2.1.2. EfficientNet

EfficientNet, a neural network designed for efficient image classification, has proven beneficial in diabetic foot diagnosis. Its scalability allows for different model sizes to accommodate computational resources and diagnostic requirements. EfficientNet's advanced feature extraction capabilities and fine-tuning on specific diabetic foot image datasets streamline the diagnostic process. Integrating EfficientNet into diagnostic systems improves efficiency and consistency, especially in time-sensitive situations. However, its effectiveness relies on carefully annotated medical image datasets, rigorous fine-tuning, and validation procedures in clinical settings.

The EfficientNet model introduces a novel approach known as the compound scaling method, which involves scaling convolutional layers to effectively extract deeper information

(Tan and Le, 2019). While previous CNN approaches primarily emphasize the search for optimal architecture layers, EfficientNet takes a different approach by examining the depth and tensor size of these layers. Therefore, EfficientNet has the potential to be integrated into the base architectures of other CNNs, such as ResNet and MobileNet. The EfficientNet framework encompasses a collection of eight distinct architectural families that are utilized within the compound scaling paradigm. The primary architect of the EfficientNet model is referred to as EfficientNetB0. The model has undergone further development, culminating in its most recent iteration, known as EfficientNetB7 (Munadi et al., 2022).

Liu et al. (2022) focus on the development of a DL method for detecting bacterial infection and inadequate blood flow (ischemia) in DFUs through image processing. The researchers utilized a dataset of DFUs and applied geometric and color image processes to augment the dataset. They then performed binary classification tasks using the EfficientNet DL model and compared its performance to other baseline models such as ResNet, Inception, and Ensemble CNN. The results showed that the EfficientNet model achieved superior performance in classifying both ischemia and infection, with accuracies of 99% and 98% respectively. This outperformed the other baseline models, which achieved accuracies of 87%. Furthermore, the EfficientNet model demonstrated a notable reduction in the time required for classifying test photos compared to the baseline models, ranging from 10% to 50%. Overall, this study provides evidence that EfficientNets can be considered a feasible DL model for accurately classifying infections and ischemia in DFUs. Basiri et al. (2022) focus on evaluating and selecting the most accurate feature extractor for the development of a DL wound detection network. The researchers utilized mAP and F1-score parameters to assess the performance of their approach using the publicly available DFU2020 dataset. They found that a hybrid approach combining the UNet architecture and the EfficientNetb3 feature extractor achieved the best results compared to the other 14 networks considered. This combination of UNet and EfficientNetb3 is deemed suitable for creating a comprehensive autonomous wound detection pipeline specifically tailored for the DFU domain. The study concludes that the EfficientNetb3 feature extractor produces the highest Intersection over Union and F1 values. Through a thorough evaluation of feature extractors and architectures, the researchers argue that an optimized combination of EfficientNetb3 and UNet for box detection has the potential to outperform the current. Liu et al. (2023) introduced a novel DL architecture called Semi-Supervised PMG EfficientNet (SS-PMG-EfficientNet), which was utilized to estimate all eight sub-scores related to PWAT. The researchers employed transfer learning techniques on the SS-PMG-EfficientNet model in order to train individual models for each of the eight PWAT sub-scores. The SS-PMG-EfficientNet architecture demonstrated strong performance in the rigorous evaluation of chronic wounds, specifically in the assessment of DFUs, vascular ulcers, pressure ulcers, and surgical wounds. The proposed approach, known as SS-PMG-EfficientNet, achieved an average classification accuracy and F1 score of approximately 90% for all 8 PWAT sub-scores. It outperformed a comprehensive set of baseline models and demonstrated a 7% improvement over the previous state-of-the-art method, without the use of data augmentation.

2.1.3. ResNet

The integration of ResNet, a DNN design known for its depth and efficacy in addressing the vanishing gradient problem, shows promise in the field of diabetic foot diagnosis. ResNet's deep layers have the ability to acquire complex details from medical pictures, making it suitable for detecting subtle patterns and abnormalities relevant to diabetic foot issues. The use of residual connections in ResNet allows for the training of deep networks with a high number of layers while maintaining the smooth transfer of information, making it suitable for medical picture processing. The depth of ResNet refers to its ability to extract pertinent characteristics from images, which is essential for accurately identifying anomalies associated with diabetic foot conditions. ResNet's adaptability in transfer learning allows for the refinement of pre-existing models using specific datasets focused on diabetic foot images, improving its ability to adapt to medical image analysis tasks with limited data. The incorporation of ResNet into diagnostic systems enhances the efficiency of automated analysis for diabetic foot pictures, accelerating the diagnostic process and enabling prompt intervention. ResNet's interpretability and visualization capabilities also enhance its usefulness in clinical settings, allowing healthcare practitioners to better understand the model's diagnostic reasoning. However, the effectiveness of ResNet in diagnosing diabetic foot relies on carefully annotated medical image datasets and rigorous processes of model fine-tuning and validation, which are necessary for its implementation in clinical environments.

The success of ResNet can be attributed to its incorporation of the identity shortcut connection, which allows for the bypassing of layers. This feature has greatly improved various computer vision applications, such as image classification, object identification, face recognition, and semantic segmentation. The inclusion of residual units and skip connections in deep networks has made training easier and information propagation more efficient. These enhancements have also reduced the number of parameters required while maintaining or improving performance in semantic segmentation tasks. Additionally, the use of average pooling instead of FC layers helps to prevent overfitting and leads to increased accuracy (Bouallal et al., 2022).

Bouallal et al. (2022) develop an automated and accurate method for segmenting diabetic foot images. The authors propose a DNN framework called Double Encoder-ResUnet that combines the advantages of residual networks and U-Net architecture. The network uses RGB color photographs and integrates thermal and color data to increase segmentation accuracy. The dataset consists of 398 pairs of RGB and thermal images, with two groups: a healthy cohort of 54 individuals and a diabetic cohort of 145 individuals. The dataset is divided into training, testing, and validation subsets. The proposed model achieves high performance in generating precise segmentations of the diabetic foot, outperforming existing techniques with an average intersection over union score of 97%. It also accurately identifies areas of the foot that are at a higher risk for ulcers, such as the toes and heels. Ahsan et al. (2023) introduce a variety of end-to-end CNN architectures, including AlexNet, VGG16/19, GoogLeNet, ResNet50.101, MobileNet, SqueezeNet, and DenseNet. These designs are used for the purpose of categorizing infections and ischemia using the benchmark dataset DFU2020. The weight is adjusted to address the issue of insufficient data and minimize computational costs. The use of affine transform methods is employed for the purpose of augmenting input data. According to the findings, the ResNet50 model demonstrates superior accuracy rates of 84.76% and 99.49% for the detection of infection and ischaemia, respectively.

2.2. Performance Assessment

The effectiveness of ML algorithms, including DL algorithms, is evaluated using performance assessment metrics. Various performance evaluation measures are used to evaluate the performance of DL models, particularly in the context of DFUs. The accurate utilization of these metrics is crucial in determining the effectiveness and optimal functioning of the model. One important component of DL models is the confusion matrix, which provides insights into the accuracy of both actual and expected classifications. The definitions of certain concepts inside the confusion matrix are as follows: in the context of DL models, when a prediction accurately identifies the positive class, it is referred to as a true positive (TP); conversely, if the prediction incorrectly identifies the positive class, it is known as a false positive (FP). In the context of DL models, when a prediction accurately identifies the negative class, it is referred to as a true negative (TN). Conversely, if the prediction incorrectly fails to identify the negative class, it is classified as a false negative (FN). The aforementioned concepts are employed in the performance assessment criteria utilized for DL models in the context of DFUs (Zhang et al., 2022). Table 1 shows the confusion matrix, and Table 2 indicates the performance assessment parameters.

Table 1. Confusion-matrix (Wang et al., 2022)

		Predicted	
		N	**P**
Actual	Negative	**TN**	**FP**
	Positive	**FN**	**TP**

N: Negative; P: Positive; TN: True negative; FN: False negative; FP: False positive; TP: True positive

- Accuracy assesses the accuracy of a model in predicting tasks. Its usefulness depends on the context and its simplicity of interpretation. However, it is vulnerable to class imbalance and may not be sufficient in situations with imbalanced class distributions or incomparable FP and FN. Therefore, accuracy should be combined with additional metrics tailored to the dataset's task objectives and features for a more nuanced evaluation.
- Sensitivity, also known as TP rate, evaluates a model's ability to accurately detect positive instances in binary classification tasks. High sensitivity reduces FN, enhancing its usefulness in medical diagnosis. However, enhancing sensitivity may lead to incorrect positive identifications. Hence, it's essential to assess the model's overall performance, considering both sensitivity and specificity, to determine its capacity to differentiate negative instances.
- Specificity is crucial for identifying negative instances in binary classification tasks, known as the TN rate. It reduces FPs and unnecessary alerts, especially in medical diagnostic tests. However, achieving greater precision may result in more FN. Therefore, it's essential to assess the model's overall performance, often in conjunction with sensitivity, to determine its capacity to accurately identify positive instances.
- Precision assesses a model's effectiveness in identifying positive instances. It is used in various tasks like search engine ranking and email spam filtering. Precision prioritizes positive predictions, reducing FP. However, it can lead to increased FN. To evaluate a model's performance, a balance between precision and recall is necessary, considering the trade-offs between these metrics.
- Recall assesses a model's ability to accurately identify positive instances in classification tasks. It represents the ratio of TP predictions to actual positive instances. Recall is especially important in medical diagnosis and quality control, where overlooking favorable occurrences can have significant consequences. However, achieving maximum recall can lead to FP, highlighting the need for a balance between recall and precision.
- The Area Under the Receiver Operating Characteristic Curve (AUC) assesses a model's ability to differentiate between classes in binary classification scenarios. It uses the ROC curve, a visual representation, to evaluate a model's discriminatory ability. AUC is particularly useful in imbalanced datasets and can withstand threshold-specific factors. However, it doesn't provide insight into the ideal threshold for specific applications, necessitating the use of precision-recall curves.
- The F1 Score combines precision and recall evaluating a model's performance in binary classification tasks. It ranges from 0 to 1, with higher values indicating better performance. It's useful in situations where high precision and recall are crucial, such as information retrieval, text categorization, and medical diagnosis. It helps balance efforts for optimal model performance.
- Average Precision (AP) is a crucial in object detection and information retrieval. It evaluates a model's ability to balance precision and recall, with higher values

indicating better performance. AP is particularly useful in situations with imbalances between positive and negative occurrences or tasks requiring high precision across different recall levels.

- Mean Average Precision (mAP) is used for object detection and classification across multiple classes. It calculates AP for each object class separately and takes the mean. It evaluates a model's precision-recall trade-offs, but its computational requirements can be significant.
- The Dice Similarity Coefficient measures the level of similarity between predicted and ground truth sets. It is crucial in applications like medical image segmentation and computer vision, where object delineation precision is crucial. However, it may not consider object size or shape differences or the distribution of errors in space. Despite these limitations, it remains a significant tool for segmentation accuracy.

Table 2. Performance assessment parameters (Zhang et al., 2022)

Accuracy	$Accuracy = \frac{TN + TP}{TN + TP + FN + FP}$
Sensitivity	$Sensitivity = \frac{TP}{(TP + FN)}$
Specificity	$Specificity = \frac{TN}{(TN + FP)}$
Precision	$Precision = \frac{TP}{FP + TP}$
Recall	$Recall = \frac{TP}{TP + FN}$
AUC	$AUC = \frac{\sum_{ins_i \in postiveclass} rank_{ins_i} - \frac{M \times (M+1)}{2}}{M \times N}$
F1-Score	$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision}$
Average precision (AP)	$AP = \frac{\sum_{q=1}^{Q} \frac{TP_i}{TP_i + FP_i}}{Q}$
Mean average precision (mAP)	$mAP = \frac{\sum_{q=1}^{Q} AveP(q)}{Q}$
Dice similarity coefficient (DSC)	$DSC = \frac{2 \times TP}{2 \times TP + FP + FN}$

2.3. Limitations and Possible Solutions

The use of DL in diabetes research has shown promising advancements, but it is important to consider qualities like robustness, reliability, and effectiveness in healthcare applications. There are several constraints that limit the implementation of DL in real-world clinical environments, including, data volume, transparency, data quality, and interpretability (Zhu et al., 2020). The major challenge is the limited availability and quality of data, particularly in obtaining comprehensive and well-annotated medical imaging datasets. Data obtained from diabetic foot patients may have imperfections due to human errors and sensor abnormalities, and acquiring empirical data can be costly and timeintensive. Data privacy policies can also make it challenging to exchange data sets among research teams. Additionally, there is a lack of transparency, making it difficult for doctors to understand the logic behind the model's output. Balanced performance and interpretability are crucial when exploring DL techniques. To overcome these limitations and ensure safe and effective integration into clinical practice, the progress of DL in identifying diabetic foot conditions requires the involvement of multiple stakeholders. The application of DL algorithms involves multiple strategies to overcome these complex challenges. These strategies include data augmentation to increase the amount and diversity of datasets, transfer learning to leverage pre-trained models and adapt them for the specific task. Rigorous clinical validation is conducted to ensure the safety and reliability of DL models in real-world clinical environments. Interdisciplinary collaboration between data scientists, healthcare practitioners, regulatory authorities, and technologists is essential for integrating technology into clinical practice.

2.3.1 Data Augmentation

Data augmentation is a crucial technique used in DL models, particularly in the field of diabetic foot diagnosis and medical imaging. It involves intentionally modifying existing data to artificially increase the size and diversity of the dataset. This is important because DL models require a large amount of training data due to the numerous parameters involved. Data augmentation is a cost-effective strategy to generate a substantial amount of data when it is not readily available. It involves various processing methods such as rotation, flipping, contrast enhancement, and changes in space and color, as well as random scaling (Cruz-Vega et al., 2020). These transformations aim to replicate real-world variations observed in medical images, such as changes in orientation, lighting conditions, image quality, and the inclusion of clinical annotations. By increasing the dataset size through data augmentation, DL models become more robust and capable of generalizing, allowing them to better handle the complexities and variations associated with diagnosing diabetic foot conditions. Additionally, data augmentation helps prevent overfitting and improves the accuracy of the diagnostic process.

Anaya-Isaza, and Zequera-Diaz (2022a) proposed the incorporation of three DL architectures using three different data augmentation approaches. A novel approach involving the use of the Fourier transform to change image amplitudes was introduced. The study utilized two CNNs and a novel approach using attention models, specifically the Transformer model. The combination of the Fourier approach and image flipping augmentation resulted in the best performance across the three networks: DFTNet, Transformer, and ResNet50v2. The networks achieved performance levels exceeding 95%, with ResNet50v2 achieving a flawless score of 100% without overfitting. The design also addressed the issue of ambiguous probabilities, with a limited number of topics having low probabilities that made classification challenging. Furthermore, the network allowed for an uncertainty threshold of up to 20%, ensuring high classification efficiency by eliminating values between 0.4% and 0.6%. Goyal et al. (2020) present a new dataset and utilize computer vision algorithms to detect infection and ischaemia in DFUs. The dataset includes ground truth labels for cases of infection and ischaemia, making it valuable for research purposes. The study introduces a new feature descriptor called the Superpixel Colour Descriptor and employs an Ensemble CNN model to improve the accuracy of identifying infection and ischaemia. The study suggests using a natural data augmentation method that focuses on the region of interest in foot images to enhance the detection of important elements. The main objective is to accurately classify ischaemia and infection through binary classification tasks. Overall, the proposed methodologies outperformed handmade ML algorithms, particularly in the classification of ischaemia. The results demonstrate that the Ensemble CNN DL algorithms achieved a classification accuracy of 90% for ischaemia and 73% for infection. Anaya-

Isaza, and Zequera-Diaz (2022b) investigate a novel discrimination coefficient based on average temperature, age, and the Temperature Coefficient Index (TCI) of the subjects' feet. This coefficient showed higher accuracy compared to TCI, with a 17% improvement. The study also explored the use of the ResNet50v2 network and data augmentation techniques to categorize participants. Twelve data augmentation techniques, including dimension reduction methods, were used to generate synthetic images. All approaches showed statistical significance in enhancing the data. A comparative analysis of the network's behavior under different training settings was conducted, including training from inception, transfer learning from the ImageNet database, and transfer learning from a thermographic database. The findings showed that data augmentation and transfer learning techniques significantly improved the performance of CNNs. The effect of transfer learning was consistent regardless of image characteristics, as long as the dataset was extensive enough to develop transferable patterns of learning. Hyun et al. (2021) offer a specialized method for synthesizing sensor-based medical time series data, with a specific focus on training models for diagnosing diabetic foot conditions. The suggested system employs statistical approaches, augmentation methods, and the NeuralProphet model to achieve its objectives while upholding medical validity. The findings of the study indicate that the synthetic time series data generated exhibit patterns and characteristics consistent with those observed in real data. In addition, their work is subjected to verification through the utilization of ML-based clustering techniques. The successful clustering of the synthetic data created by their suggested system serves as empirical evidence to support the assertion that their system can achieve its intended objectives.

3. Results and Discussion

This paper critically evaluates several DL methods employed in the diagnosis of diabetic foot. Each reviewed study, which includes a DL method for diabetic foot diagnosis, utilizes diverse methodologies that offer both advantages and disadvantages. To comprehensively assess these advantages and disadvantages, it is essential to simultaneously investigate various parameters. The evaluation of performance can only be accomplished with this strategy. Based on this explanation, Table 3 simultaneously assessing these characteristics, it is possible to draw conclusions regarding the efficacy of DL approaches. Subsequently, the selection of a method that aligns with the work’s objectives allows the achievement of the most favorable outcomes.

4. Conclusion

According to the International Diabetes Federation, the prevalence of diabetes among adults worldwide is projected to increase to 700 million by 2045. Additionally, approximately one-third of individuals diagnosed with diabetes will develop a diabetic foot condition, with a lifetime probability. The issue at hand carries a substantial probability of leading to the requirement of amputation. Therefore, the ability to identify diabetic foot problems accurately is crucial for timely intervention.

In this study, focused on the use of DL methodologies for the diagnosis of diabetic foot conditions. These methodologies utilize medical images as datasets which obtained various sources such as diverse medical images like X-rays, CT scans, MRI scans, and PET scans. Additionally, they include various medical resources such as electronic medical records, genomics data, bioinformatics data, and drug response data. However, datasets are often accompanied by challenges. To address these challenges related to limited data, overfitting, and imbalanced class distribution, researchers have employed data augmentation methods, regularization techniques, and ensemble learning. Also, clinical validation is crucial to ensure the reliability and effectiveness of DL models in real-world scenarios. Furthermore, ethical, and legal frameworks are in place to ensure compliance with patient consent and data privacy regulations.

Table 3. Examination of Related Works

Study	Aim	Network Structure	Data Set	Limitations	Best Results	Data Set Information
Yap et al., (2021b)	DFU detection	Faster R-CNN	DFUC2020	Not mentioned	mAP=0.6940, F1-Score=0.7434.	2, 000 images, with 640×480 pixels.
Ahsan et al., (2023)	DFU infection and ischemia categorization	ResNet50	DFU2020	Not mentioned	Accuracy=99.49%, Sensitivity=99.59%, Specificity=99.39%, Precision=99.39%, F-Score=99.49% AUC=99.96%	1459 images with sizes ranging from 1600 × 1200 and 3648 × 2736 pixels.
Liu et al., (2022)	DFU infection and ischemia detection	EfficientNet	DFUC2021	1) High inter-class similarity and intra-class variations in DFU images; (2) Variable and unstandardized DFU data set due to the unstable camera conditions each time the images were captured (3) Lack of differential demographic information of patients.	Accuracy=99% in ischemia classification and Accuracy=98% in infection classification	15760 images with 224 × 224 pixels.
Yap et al., 2021a)	DFU infection and ischaemia classification	Efficient- NetB0	DFUC2021	The detection of infection and detection of co-occurrences of both ischaemia and infection.	Precision=0.57, Recall=0.62 Accuracy= 97.9%.	15,683 images with, 224X224 pixels.
Prabhu and Verma, (2021)	Distinguishing healthy skin and DFU class	Proposed app.		The time spent on classification tasks, such as collecting and labeling images, and distinguishing between healthy and abnormal skin, is increasing.	Accuracy= 97.9%.	122 images
Toofanee et al., (2023)	DFU infection and ischemia categorization	Siamese Neural Network (SNN)	DFUC2021	Quality of the dataset	Macro F1-score= 0.6455	11,000 images

Table 3 Continuation: Examination of Related Works

Study	Aim	Network Structure	Data Set	Limitations	Best Results	Data Set Information
Goyal et al., (2020)	DFU infection and ischaemia classification	Ensemble CNN	Lancashire Teaching Hospitals	There are significant visual differences between classes, as well as similarities between classes.	Accuracy=90% in ischaemia classification and Accuracy=73% in infection classification.	1459 images, size varies between 1600×1200 and 3648×2736 pixels.
Alzubaidi et al., (2020)	Distinguishing the healthy skin and DFU class	DFU_QUTNet	Nasiriyah Hospital	Not mentioned	F1-score=94.5%	754 images
Goyal et al., (2018)	Distinguishing the healthy skin and DFU class	DFUNet	Lancashire Teaching Hospitals	It costs a lot to diagnose, treat, and care for people with DFU in the long run.	Sensitivity=0.934, F-measure= 0.939, AUC=0.962	397 images with 256×256 pixels.
Rania et al., (2020)	DFU segmentation	U-Net	Centre Hospitalier Regional d'Orleans	Not mentioned	Dice=97.25, Accuracy=94.96	92 images
Oliveira et al., (2021)	DFU detection	Faster R-CNN	DFUC 2020	Not mentioned	Precision=91.4%, F1-score=94.8%	2000 images with 640x480 pixels.
Cao et al., (2023)	DFU semantic segmentation	Mask R-CNN	Diabetic Foot Prevention and Treatment Center of Xiangya Hospital, and DFUC2020	Not mentioned	Specificity=99.50%, Sensitivity=70.62%, Precision = 84.56%, Mean Average Precision=85.70%	3000 images with 512x512 pixels

Collaboration between data scientists, healthcare professionals, and technologists, is essential for advancing knowledge and innovation in healthcare. DL methodologies have the potential to transform the diagnosis of diabetic foot conditions, improving precision, efficiency, and availability in clinical settings. The future of DL algorithms in diabetic foot diagnosis holds great promise, with several notable advancements expected. These advancements will be driven by the continuous improvement of data sources, model architectures, and interpretability techniques.

Numerous scientific studies have been conducted on this topic. However, these studies mostly focus on only a few types of deep learning methods for a few specific types of diseases. Typically, studies utilize one specific data set. In this context, the comprehensive analysis of existing research within the literature has significant importance in terms of offering a holistic understanding of such a critical problem. This research has significant value as it conducts a comprehensive analysis of deep learning methods related to differential diagnosis in existing papers, in addition to demonstrating the diverse outcomes given by various parameters. Examining different data sets, deep learning methods, and diabetic foot problems from a holistic standpoint would provide a valuable contribution to the literature.

Authors Contributions

M. Çakır: Investigation, Formal Analysis, Writing, Editing; **H. Canbolat:** Resources, Review & Editing; **G. Tulum:** Resources, Review & Editing.

References

Ahsan, M., Naz, S., Ahmad, R., Ehsan, H., and Sikandar, A. (2023). A deep learning approach for diabetic foot ulcer classification and recognition. *Information*, 14(1), 36. https://doi.org/10.3390/info14010036

Alzubaidi, L., Fadhel, M. A., Oleiwi, S. R., Al-Shamma, O., and Zhang, J. (2020). DFU_QUTNet: diabetic foot ulcer classification using novel deep convolutional neural network. *Multimedia Tools and Applicat*ions, 79(21-22), 15655-15677. https://doi.org/10.1007/s11042-019-07820-w

Amin, J., Sharif, M., Anjum, M. A., Khan, H. U., Malik, M. S. A., and Kadry, S. (2020). An integrated design for classification and localization of diabetic foot ulcer based on CNN and YOLOv2-DFU models. *IEEE Access*, 8, 228586-228597. https://doi.org/10.1109/ACCESS.2020.3045732

Anaya-Isaza, A., and Zequera-Diaz, M. (2022a). Detection of diabetes mellitus with deep learning and data augmentation techniques on foot thermography. *IEEE Access, 10*, 59564-59591. https://doi.org/10.1109/ACCESS.2022.3180036

Anaya-Isaza, A., and Zequera-Diaz, M. (2022b). Fourier transform-based data augmentation in deep learning for diabetic foot thermograph classification. *Biocybernetics and Biomedical Engineering, 42*(2), 437-452. https://doi.org/10.1016/j.bbe.2022.03.001

Basiri, R., Popovic, M. R., and Khan, S. S. (2022). Domain-Specific Deep Learning Feature Extractor for Diabetic Foot Ulcer Detection. *IEEE International Conference on Data Mining Workshops (ICDMW)*, 2022, Orlando, FL, USA. https://doi.org/10.1109/ICDMW58026.2022.00041

Belsti, Y., Akalu, Y., and Animut, Y. (2020). Attitude, practice and its associated factors towards Diabetes complications among type 2 diabetic patients at Addis Zemen District hospital, Northwest Ethiopia. *BMC Public Health, 20*, 1-11. https://doi.org/10.1186/s12889-020-08953-6

Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks, 5*(2), 157-166. https://doi.org/10.1109/72.279181

Bouallal, D., Douzi, H., and Harba, R. (2022). Diabetic foot thermal image segmentation using Double Encoder-ResUnet (DE-ResUnet). *Journal of Medical Engineering and Technology, 46*(5), 378-392. https://doi.org/10.1080/03091902.2022.2077997

Cao, C., Qiu, Y., Wang, Z., Ou, J., Wang, J., Hounye, A. H., . . . Zhang, J. (2023). Nested segmentation and multi-level classification of diabetic foot ulcer based on mask R-CNN. *Multimedia Tools and Applications, 82*(12), 18887-18906. https://doi.org/10.1007/s11042-022-14101-6

Chamberlain, R. C., Fleetwood, K., Wild, S. H., Colhoun, H. M., Lindsay, R. S., Petrie, J. R., . . . Sattar, N. (2022). Foot ulcer and risk of lower limb amputation or death in people with diabetes: a national population-based retrospective cohort study. *Diabetes Care, 45*(1), 83-91.

https://doi.org/10.2337/dc21-1596

Chan, H.-P., Samala, R. K., Hadjiiski, L. M., and Zhou, C. (2020). Deep learning in medical image analysis. *Deep Learning in Medical Image Analysis: Challenges and Applications*, 3-21. https://doi.org/10.1007/978-3-030-33128-3_1

Cruz-Vega, I., Hernandez-Contreras, D., Peregrina-Barreto, H., Rangel-Magdaleno, J. d. J., and Ramirez-Cortes, J. M. (2020). Deep learning classification for diabetic foot thermograms. *Sensors, 20*(6), 1762. https://doi.org/10.3390/s20061762

Oliveira, A. L., de Carvalho, A. B., and Dantas, D. O. (2021). Faster R-CNN Approach for Diabetic Foot Ulcer Detection. *16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP volume 4: 677-684p),* 2021, Vienna, Austria. https://doi.org/10.5220/0010255506770684

Das, S. K., Roy, P., and Mishra, A. K. (2022). DFU_SPNet: A stacked parallel convolution layers based CNN to improve Diabetic Foot Ulcer classification. *ICT Express, 8*(2), 271-275. https://doi.org/10.1016/j.icte.2021.08.022

Das, S. K., Roy, P., Singh, P., Diwakar, M., Singh, V., Maurya, A., . . . Kim, J. (2023). Diabetic foot ulcer identification: A Review. *Diagnostics, 13*(12), 1998. https://doi.org/10.3390/diagnostics13121998

Ferreira, A. C. B. H., Ferreira, D. D., Oliveira, H. C., de Resende, I. C., Anjos, A., and de Moraes Lopes, M. H. B. (2020). Competitive neural layer-based method to identify people with high risk for diabetic foot. *Computers in biology and medicine, 120*, 103744. https://doi.org/10.1016/j.compbiomed.2020.103744

Gamage, C., Wijesinghe, I., and Perera, I. (2019). Automatic scoring of diabetic foot ulcers through deep CNN based feature extraction with low rank matrix factorization. *IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*, 2019, Athens, Greece, 2019, pp. https://doi.org/10.1109/BIBE.2019.00069

Goyal, M., Reeves, N. D., Davison, A. K., Rajbhandari, S., Spragg, J., and Yap, M. H. (2018). Dfunet: Convolutional neural networks for diabetic foot ulcer classification. *IEEE Transactions on Emerging Topics in Computational Intelligence,* 4(5), 728-739. https://doi.org/10.1109/TETCI.2018.2866254

Goyal, M., Reeves, N. D., Rajbhandari, S., Ahmad, N., Wang, C., and Yap, M. H. (2020). Recognition of ischaemia and infection in diabetic foot ulcers: Dataset and techniques. *Computers in Biology and Medicine, 117*, 103616. https://doi.org/10.1016/j.compbiomed.2020.103616

Harahap, M., Anjelli, S. K., Sinaga, W. A. M., Alward, R., Manawan, J. F. W., and Husein, A. M. (2022). Classification of diabetic foot ulcer using convolutional neural network (CNN) in diabetic patients. *Jurnal Infotel,* 14(3), 196-202. https://doi.org/10.20895/infotel.v14i3.796

Hyun, J., Lee, Y., Son, H. M., Lee, S. H., Pham, V., Park, J. U., and Chung, T.-M. (2021). Synthetic data generation system for AI-based diabetic foot diagnosis. *SN Computer Science,* 2(5), 345. https://doi.org/10.1007/s42979-021-00667-9

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature,* 521(7553), 436-444. https://doi.org/10.1038/nature14539

Liu, Z., Agu, E., Pedersen, P., Lindsay, C., Tulu, B., and Strong, D. (2023). Chronic wound ımage augmentation and assessment using semi-supervised progressive multi-granularity EfficientNet. *IEEE Open Journal of Engineering in Medicine and Biology, (Early Access)*. https://doi.org/10.1109/OJEMB.2023.3248307

Liu, Z., John, J., and Agu, E. (2022). Diabetic Foot Ulcer Ischemia and Infection Classification Using EfficientNet Deep Learning Models. *IEEE Open Journal of Engineering in Medicine and Biology,* 3, 189-201. https://doi.org/10.1109/OJEMB.2022.3219725

Maltese, G., Koufakis, T., Kotsa, K., Basile, G., and Siow, R. (2023). Mediterranean diet, type 2 diabetes prevention and healthy ageing: do we need more evidence? *Diabetes research and clinical practice (In Press, Journal Pre-proof),*110928. https://doi.org/10.1016/j.diabres.2023.110928

Min, S., Lee, B., and Yoon, S. (2017). Deep learning in bioinformatics. *Briefings in bioinformatics, 18*(5), 851-869. https://doi.org/10.1093/bib/bbw068

Munadi, K., Saddami, K., Oktiana, M., Roslidar, R., Muchtar, K., Melinda, M., . . . Arnia, F. (2022). A deep learning method for early detection of diabetic foot using decision fusion and thermal images. *Applied Sciences, 12*(15), 7524. https://doi.org/10.3390/app12157524

Ogurtsova, K., Guariguata, L., Barengo, N. C., Ruiz, P. L.-D., Sacre, J. W., Karuranga, S., . . . Magliano, D. J. (2022). IDF diabetes Atlas: Global estimates of undiagnosed diabetes in adults for 2021. *Diabetes research and clinical practice, 183*, 109118. https://doi.org/10.1016/j.diabres.2021.109118

Prabhu, M. S., and Verma, S. (2021). A Deep Learning framework and its Implementation for Diabetic Foot Ulcer Classification. *9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future*

Directions)(ICRITO), 2021, Noida, India. https://doi.org/10.1109/ICRITO51393.2021.9596380

Rania, N., Douzi, H., Yves, L., Sylvie, T. (2020). Semantic Segmentation of Diabetic Foot Ulcer Images: Dealing with Small Dataset in DL Approaches. In: El Moataz, A., Mammass, D., Mansouri, A., Nouboud, F. (eds) Image and Signal Processing. ICISP 2020. Lecture Notes in Computer Science(), vol 12119. https://doi.org/10.1007/978-3-030-51935-3_17

Reardon, R., Simring, D., Kim, B., Mortensen, J., Williams, D., and Leslie, A. (2020). The diabetic foot ulcer. *Australian Journal of General Practice, 49*(5), 250-255. https://doi.org/10.31128/AJGP-11-19-5161

Reyes-Luévano, J., Guerrero-Viramontes, J., Romo-Andrade, J. R., and Funes-Gallanzi, M. (2023). DFU_VIRNet: A novel Visible-InfraRed CNN to improve diabetic foot ulcer classification and early detection of ulcer risk zones. *Biomedical Signal Processing and Control,* 86, 105341. https://doi.org/10.1016/j.bspc.2023.105341

Roback, K. (2010). An overview of temperature monitoring devices for early detection of diabetic foot disorders. *Expert review of medical devices, 7*(5), 711-718. https://doi.org/10.1586/erd.10.35

Sharma, N., Mirza, S., Rastogi, A., and Mahapatra, P. K. (2023). Utilizing Mask R-CNN for automated evaluation of diabetic foot ulcer healing trajectories: A novel approach. *Traitement du Signal,* 40(4). https://doi.org/10.18280/ts.400428

Tan, M., and Le, Q. (2019). *Efficientnet: Rethinking model scaling for convolutional neural networks.* Paper presented at the International conference on machine learning. https://doi.org/10.48550/arXiv.1905.11946

Toofanee, M. S. A., Dowlut, S., Hamroun, M., Tamine, K., Duong, A. K., Petit, V., and Sauveron, D. (2023). DFU-Helper: An innovative framework for longitudinal diabetic foot ulcer diseases evaluation using deep learning. *Applied Sciences, 13*(18), 10310. https://doi.org/10.3390/app131810310

Wang, Y., Jia, Y., Tian, Y., and Xiao, J. (2022). Deep reinforcement learning with the confusion-matrix-based dynamic reward function for customer credit scoring. *Expert Systems with Applications, 200*, 117013. https://doi.org/10.1016/j.eswa.2022.117013

Yap, M. H., Cassidy, B., Pappachan, J. M., O'Shea, C., Gillespie, D., and Reeves, N. D. (2021a). Analysis towards classification of infection and ischaemia of diabetic foot ulcers. *IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, 2021, Athens, Greece. https://doi.org/10.1109/BHI50953.2021.9508563

Yap, M. H., Hachiuma, R., Alavi, A., Brüngel, R., Cassidy, B., Goyal, M.,. Huang, X. (2021b). Deep learning in diabetic foot ulcers detection: A comprehensive evaluation. *Computers in biology and medicine, 135*, 104596. https://doi.org/10.1016/j.compbiomed.2021.104596

Zhang, J., Qiu, Y., Peng, L., Zhou, Q., Wang, Z., and Qi, M. (2022). A comprehensive review of methods based on deep learning for diabetes-related foot ulcers. *Frontiers in Endocrinology, 13*, 945020. https://doi.org/10.3389/fendo.2022.945020

Zhu, T., Li, K., Herrero, P., and Georgiou, P. (2020). Deep learning for diabetes: a systematic review. *IEEE Journal of Biomedical and Health Informatics, 25*(7), 2744-2757. https://doi.org/10.1109/JBHI.2020.3040225