International Journal of Computers and Informatics

Journal Homepage: https://www.ijci.zu.edu.eg

Int. j. Comp. Info. Vol. 7 (2025) 126–155

Paper Type: Original Article

Computer-Aided Detection of Diabetic Foot via Infrared Imaging using Machine and Deep Learning Approaches: A Survey

A. Sharaf Eldin [1,2], Asmaa S. Ahmoud [3,*] and Hanaa M. Hamza [4]

[1] Department of Information and Decision Support Systems, Faculty of Information Technology and Computer Science, Sinai University, Arish, 16020, Egypt; ahmed.sharaf@su.edu.eg.

[2] Department of Information Systems, Faculty of Computers and AI, Helwan University, Cairo, 11795, Egypt.

[3] Department of Information Technology, Faculty of Information Technology and Computer Science, Sinai University, Arish, 16020, Egypt; asmaa.salman@su.edu.eg.

[4] Department of Information Technology, Faculty of Computers and Informatics, Zagazig University, Zagazig, 44519, Egypt; hmkamal@fci.zu.edu.eg.

Received: 07 Jan 2025 **Revised:** 02 Apr 2025 **Accepted:** 27 Jun 2025 **Published:** 29 Jun 2025

Abstract

Diabetes mellitus (DM) is a major health problem and the most prevalent worldwide. Diabetic foot (DF) is the most common complication of DM, which can lead to death, amputation, and plantar ulcers. Early detection of these complications protects the diabetic patient from dangerous stages that lead to amputation. This study explores the effectiveness of computer-aided diagnostic systems, particularly those that utilize the power of artificial intelligence (AI) and deep learning (DL). AI and DL may provide promising means of early detection and diagnosis of DF complications. In addition, these have the potential to revolutionize patient care by providing tools that can analyze complex medical data with remarkable accuracy and speed. Using thermal imaging is an innovative approach that has recently gained more attention. Infrared thermal imaging captures heat emanating from the body, providing a non-invasive way to detect abnormal plantar temperatures that indicate underlying inflammation or infection. Machine learning (ML) and DL classification techniques improve the effectiveness of computer-aided detection (CAD) of DF. By training algorithms on huge data sets, these systems can learn to identify patterns and anomalies that otherwise would elude human detection. This study explores several ML techniques, with a particular emphasis on DL classification to accurately identify the feet of diabetic patients. The findings from this research will contribute to future studies aimed at improving detection processes and helping medical professionals deliver timely and effective care to their patients. By automating the initial screening process, healthcare providers can prioritize patients who need immediate attention. This will improve resource allocation and potentially reduce the incidence of serious complications such as ulcers and amputations.

Keywords: Diabetic Foot; Infrared Thermal Images; Machine Learning; Deep Learning; CNN.

1 | Introduction

Every year, more than one million diabetic individuals have their feet amputated. This is due to a lack of early ulcer detection and appropriate treatment by specialists. According to World Health Organization (WHO) reports, diabetes has become more common around the world. This affects approximately 422 million people worldwide [1]. Furthermore, it claims the lives of roughly 1.6 million individuals annually. According to the WHO, approximately 10% of pregnant women suffer from gestational diabetes. Factors such as DM, obesity, hypertension, and hyperlipidemia are influencing Egypt's national morbidity and mortality rates, according to mounting evidence. It accounts for approximately 26% of all deaths caused by chronic conditions [2].

Corresponding Author: asmaa.salman@su.edu.eg

Diabetic mellitus (DM) is a serious chronic condition that affects many people with high blood sugar levels, leading to complications such as damage to blood vessels, nerves, and eyes, which may result in ulcers. The main cause of these ulcers is a lack of proper blood circulation and nerve damage [3]. There are two types of diabetes. Type 1 diabetes occurs when the pancreas does not produce any insulin for the body, causing many problems. Lifestyle factors such as being overweight, poor diet, and lack of physical activity commonly cause type 2 diabetes. Type 2 diabetes, also known as non-insulin-dependent or adult-onset diabetes, impacts over 95% of individuals with the condition. The main factors contributing to this type of diabetes are elevated body weight and inactivity. Gestational diabetes is Hyperglycemia during pregnancy occurs when blood glucose levels are above normal. Gestational diabetes occurs during pregnancy [4].

Deep learning (DL) methods enable computers to learn, observe, and understand by gathering and analyzing massive amounts of data. Diabetic foot (DF), a disease prevalent in many societies, continues to require extensive research due to its serious nature [5]. Doctors typically manually identify this disease by examining its image or the patient's foot. Image processing techniques create an automated system for DF screening, heavily relying on the extraction of pertinent features for the classification of diabetes illnesses [6].

Classification systems play a central role in the application of artificial intelligence (AI) to medical image analysis [8]. In this context, image processing enables the extraction of relevant features from 2D and 3D datasets, supporting diagnostic decisions and treatment planning. AI models typically operate in two phases: training and testing [9-10]. During training, the model learns from labeled images, identifying patterns and constructing decision rules based on key features such as shape, texture, or pressure points. These features help distinguish between clinical outcomes, such as identifying the risk of diabetic foot ulcers (DFUs). Feature selection techniques are used to retain only the most relevant data, improving the model's accuracy and efficiency. Once trained, the model enters the testing phase, applying learned knowledge to predict outcomes for new, unseen images. Through this process, classification systems enable accurate and automated interpretation of medical images, improving early detection and patient care in DFU cases.

An accurate treatment for DF primarily depends on the correct diagnosis, often aided by advanced imaging techniques. The more advanced the image quality, the higher the accuracy. These include X-rays, infrared shafts, ultrasound, magnetic resonance imaging (MRI), necropsies, and thermography. When it comes to dealing with images, they can be either 2D, which uses vertical and perpendicular dimensions (X and Y), or 3D, which adds depth (Z) [11].

X-rays are electromagnetic radiation that can penetrate solid objects, such as the human body, to create internal images for diagnosing and treating medical conditions. Medical imaging uses invisible beams of light, known as infrared shafts, to detect heat and identify tumors, fractures, and other abnormalities [12]. Ultrasound commonly diagnoses pregnancy and other medical conditions by using sound waves to create internal body images [13]. MRI uses magnetic fields and radio waves for detailed internal body images, aiding in the diagnosis of several medical conditions [14]. Necropsies, or postmortem examinations, determine the cause of death or other medical information. Animals can also undergo this procedure to diagnose medical conditions. Using the IR region of the electromagnetic spectrum to find emitted energy, infrared thermography (IR) converts IR radiation (heat) into a visible image [15]. In the medical field, IR has emerged as a recent technique for examining diabetic feet by detecting thermal alterations in the affected areas [16–17].

A thermal camera captures infrared thermal imaging, creating an image of the infrared radiation an object emits. An infrared image represents each pixel as a thermal position [18–19]. IR technology has gained interest in research due to its non-invasive nature and lack of radiation exposure, making it a safe option for medical diagnosis. It is also more accurate and can detect subtle temperature changes that may indicate disease [16]. The accuracy of surface temperature measurements in two-dimensional images decreases, especially when dealing with complex objects such as human body parts, frequently resulting in false-positive findings. Therefore, it is not advisable to use 2D thermal imaging as a medical diagnostic tool [20–21].

The most recent study on DF detection dates back to 2025 [22]. focusing on the potential of machine learning (ML) techniques to enhance diagnostic accuracy and improve treatment approaches. However, it lacked a comprehensive comparative analysis of commonly used AI methods and did not provide sufficient insights into their practical applicability across different datasets or clinical settings. Moreover, most existing studies have focused on individual models or specific imaging types without evaluating the broader landscape of both ML and DL techniques. Therefore, the motivation for this survey is to address these limitations by offering a broader and more comparative perspective:

- This study provides a summary of numerous recent studies on the identification of diabetic feet, specifically focusing on DL techniques to enhance their accuracy in detecting the disease.
- This study also suggests future research directions by highlighting the research gaps and challenges of existing studies.

The paper is structured in the following manner: Section two provides an explanation of the classification systems, while sections three and four offer an overview of ML and DL, respectively. Section five presents performance parameters; section six presents relevant previous research conducted in this field; and section seven outlines the main new research directions. Lastly, section eight contains the conclusions.

2 | Machine Learning (ML)

Traditional (shallow) ML plays a crucial role in automating medical diagnostics, including the early detection of DF. By learning from handcrafted features in thermal or clinical images, ML algorithms can accurately classify foot conditions based on thermal or clinical images, thus improving patient outcomes and supporting medical decision-making [23].

2.1 | Categories of Machine Learning

ML approaches can be categorized into four broad groups; each serving distinct purposes in the realm of data analysis and decision-making. These are supervised, unsupervised, semi-supervised, and reinforcement learning [23], as shown in Table 1.

Table 1: Machine Learning Approaches in DF Detection.

Shallow Type	Definition	Applications in DF detection
Supervised [24]	Uses labeled datasets to train models that predict outcomes based on input features.	Classifying thermogram images as ulcer-prone or healthy.
Unsupervised [25]	Analyzes unlabeled data to identify patterns or clusters.	Grouping foot images based on hidden patterns in temperature.
Semi-supervised [26]	Combines a small amount of labeled data with a large set of unlabeled data.	Enhancing model accuracy using a few labeled foot images.
Reinforcement [27]	Trains models through reward-based learning in dynamic environments.	Optimizing pressure-relief treatment plans via interactive feedback.

2.2 | Machine Learning Techniques

The common ML techniques include decision trees (DT), logistic regression (LR), support vector machines (SVM), K-nearest neighbor (KNN), Naive Bayes (NB) [28]. The following section outlines commonly applied ML techniques in DF classification.

- **Decision Trees (DT):** Decision Trees split data based on features to form branches leading to classification outcomes. In diabetic foot studies, DTs can identify key risk indicators like skin temperature or previous ulcer history [29-31].

- **Support Vector Machines (SVM):** SVMs create optimal boundaries between classes in high-dimensional data. For DF diagnosis, SVMs are used to distinguish between affected and unaffected thermographic patterns [32-34].
- **K-Nearest Neighbor (KNN):** KNN classifies new instances based on the majority label of the closest examples. In DF, it can compare foot images to known ulcer cases to determine risk levels [35-36].
- **Logistic Regression (LR):** LR estimates the probability of an outcome (e.g., presence of DF) using a logistic function. It is effective when features such as BMI, glucose levels, and image-based scores are available [37-38].
- **Naïve Bayes (NB):** NB classifiers use probabilistic reasoning based on Bayes' theorem. It's useful in binary classification, such as determining whether a patient is likely or unlikely to develop ulcers based on symptom presence [39-40].

3 | Deep Learning (DL)

DL is a branch of ML that employs multiple layers to extract both high-level and low-level data from inputs such as images, numerical values, and categorical values. As shown in Figure 1, DL differs from ML in that DL combines feature extraction and classification into a single layer, whereas ML uses separate layers for these processes. These days, artificial neural networks (ANNs), multi-layer perceptrons (MLPs), recurrent neural networks (RNNs), convolutional neural networks (CNNs), generative adversarial networks (GANs), transformers, autoencoders, and deep Q networks are used to build most DL models. These models can be combined with other DL models such as generative models, deep belief networks, and the Boltzmann machine [41].

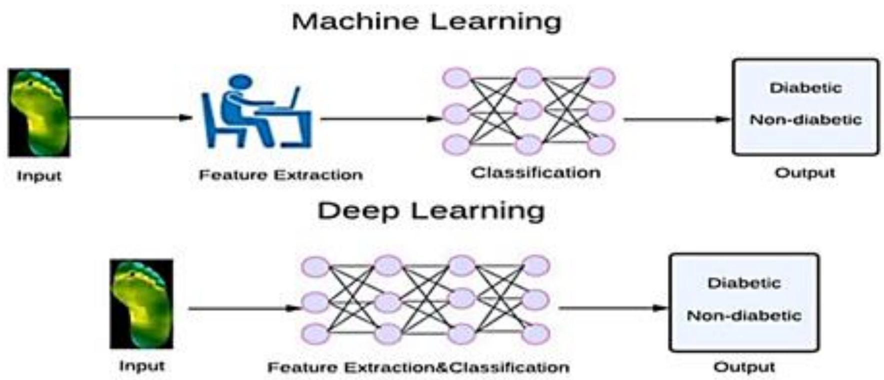

Figure 1. The difference between ML and DL [42].

Each level of DL captures new features and transforms the data to the next level. For example, in face recognition applications, the initial layers detect edges, then the subsequent layers identify features such as noses and eyes, while the third layer recognizes the entire face. Over the years, DL has shown immense potential in medical fields such as pathology, radiology, and thermography for disease diagnosis. It also has practical applications in analyzing molecular states and determining disease progression or therapy sensitivity, areas often overlooked by human investigation [43].

There are several types of neural networks that form the foundation for most pre-trained models in DL:

3.1 | Artificial Neural Networks (ANNs)

An ANN is a computational model that mimics the function of nerve cells in the human brain. It utilizes learning algorithms that allow it to adapt, perceive, and acquire knowledge independently, including MLPs, RNNs, and basic feedback models. Consequently, it is a valuable instrument for simulating non-linear statistical data [44].

One of the primary advantages of an ANN is its ability to learn by observing datasets, making them effective for approximating functions and distributions to find optimal solutions [44]. ANNs store data across the entire network, can function with limited information, and have the numerical strength to perform multiple tasks simultaneously. However, large neural networks require significant processing time [45].

3.1.1| Multi-Layer Perceptron (MLP)

Multiple layers of neurons, or nodes, connected in a feed-forward manner make up an MLP, a type of ANN. We use it for supervised learning tasks like classification and regression. It consists of an input layer, one or more hidden layers, and an output layer. Each layer connects neurons with weights that indicate the strength of these connections. The input layer receives data and passes it to the hidden layers, which process the data using activation functions such as sigmoid or Rectified Linear Unit (ReLU). The output layer produces a result based on the inputs it has received from the hidden layers [46]. Backpropagation trains MLPs, adjusting weights based on training errors to enhance performance. Applications such as image recognition, natural language processing, and time series forecasting utilize them [47].

3.1.2| Recurrent Neural Networks (RNNs)

An RNN contains loops that the network uses to store information. In other words, RNNs use their reasoning based on past experiences to predict future events. Machine translation is a typical use of RNNs [48].

The RNN applies identical actions to all inputs and hidden layers, resulting in an output that employs the same parameters for each input. This approach reduces the complexity of the parameter set compared to other neural networks. The drawbacks of RNN are that it performs slow and complex training operations, making it difficult to process longer sequences [49].

3.2| Advanced Deep Learning Models

We will discuss the most advanced models, including CNNs, GANs, Transformers, Autoencoders, and Deep Q Networks (DQNs), in this section. These models represent the advanced and specialized aspects of DL, including tasks such as image recognition, generative modeling, and reinforcement learning.

3.2.1| Convolution Neural Networks (CNNs)

CNNs specialize in handling structured data sets such as image arrays and are highly effective for visual tasks such as image classification, object detection, and segmentation [50]. Additionally, it has experience using natural language processing to categorize texts. The patterns in the input image, such as lines, gradients, circles, or even eyes and faces, are very well recognized by CNNs. CNNs do not require any prior image processing, in contrast to older computer vision algorithms. Image analysis applications such as segmentation, object detection, and image recognition primarily utilize CNN [51]. The CNN structure consists of three components: input, hidden layers, and output, with the number of hidden layers varying depending on the design. Hidden layers serve as intermediate levels in feedforward networks [52]. Figure 2 illustrates the general architecture of a CNN model.

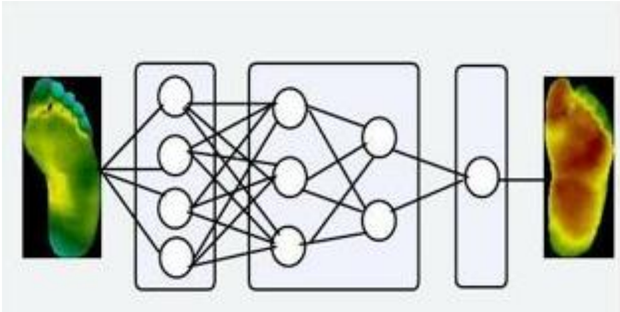

Figure 2. The general architecture of a CNN model.

Since their introduction in 1989, many CNN models have excelled in disease diagnosis over the past 30 years. Figure 3 shows the most commonly used CNN models.

A CNN consists of convolution layers, pooling layers, and fully connected layers. Its goal is to acquire spatial hierarchies of features using a backpropagation algorithm autonomously and adaptively [53]. CNN contains hidden layers and convolutional layers, which form the foundation of ConvNets. The convolutional layer receives the input volume and performs a mathematical scalar product with the feature array (filter), producing feature maps as output [54].

Employing CNNs has many advantages: CNN's weight-sharing capability reduces the number of trainable parameters, improving generalization and preventing overfitting. By learning feature extraction and classification layers together, CNNs produce highly organized outputs based on extracted features. They are easier to implement on a large scale compared to other neural networks, and they excel at image classification. On the other hand, CNN does not encode the position and orientation of an object [55].

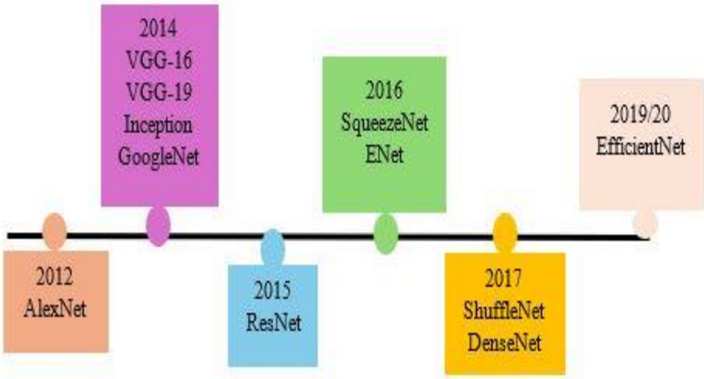

Figure 3. Widely used CNN models.

3.2.2| Generative Adversarial Networks (GANs)

GANs, introduced in 2014 by Ian Goodfellow and colleagues, are advanced DL models aiming to generate high-quality synthetic data derived from real data. This is achieved through a distinctive training method involving two neural networks: the generator and the discriminator. The generator produces synthetic data from random noise, and the discriminator distinguishes between real and synthetic data. Through iterative training, the generator improves the realism of its data, while the discriminator becomes better at identifying fake data [56].

Throughout training, the discriminator improves its ability to differentiate between real and synthetic data, creating a feedback loop. Enhancements in the generator prompt adaptations in the discriminator and vice versa, resulting in continuous improvement for both producing realistic data and accurately classifying samples [57–58].

3.2.3| Transformer

Transformer is a unique DL infrastructure that executes tasks sequentially. Many fields, including computer vision and speech processing, use it. This is due to its efficiency in managing large data sets and long-term dependencies [59].

Key features empower the Transformer's performance, with the self-attention mechanism being central. It evaluates word importance in sequence, capturing relationships regardless of position. Multi-head attention enhances this by using multiple heads simultaneously, each learning distinct weights. To address the lack of positional information, positional encoding provides sequential order details. The encoder-decoder architecture includes multiple layers, with sub-layers featuring multi-head self-attention and position-wise feedforward networks. Layer normalization and residual connections ensure stable training. Each layer is equipped with an independent feedforward network that operates on each sequence position. Attention masks are used during training to keep track of positional information, and the model tries to reduce task-specific losses like cross-entropy for machine translation. This makes it better at handling complex sequential relationships [60].

3.2.4| Autoencoders

Autoencoders, pivotal in unsupervised learning, are essential for tasks of dimensionality reduction, feature learning, and generative modeling. These neural networks create a concise representation of input data by encoding it into a lower-dimensional space. Autoencoders, pivotal in unsupervised learning, are essential for tasks of dimensionality reduction, feature learning, and generative modeling. They work by encoding input data into a lower-dimensional space, then reconstructing it. Autoencoders include key components: an encoder, a decoder, and a latent space representing condensed input information. The goal during training is to minimize the difference between the original and reconstructed data, typically using mean squared error (MSE). Training involves encoding the input, decoding it, and updating model parameters through backpropagation with optimization techniques like stochastic gradient descent (SGD) [61].

Standard autoencoders have problems like overfitting, not regularizing enough, not being able to manage sparse or noisy input well, not being able to control the latent space, and deterministic encoding. There are several types of autoencoders available: contractive autoencoders with regularization, sparse autoencoders that promote sparsity, incomplete autoencoders for data compression, and denoising autoencoders that remove noise from images. These problems can be solved by variational autoencoders (VAEs), which show hidden characteristics as a probability distribution over a continuous latent space that makes sampling easier [62-63].

3.2.5| Deep Q Networks (DQNs)

DQNs, a robust paradigm in reinforcement learning, excel in scenarios where an agent interacts with an environment to maximize cumulative rewards. Key components include Q-Learning, experience replacing for stability and sample efficiency, and periodic updates of the target Q-network. Epsilon-Greedy Exploration balances exploration and exploitation. The training process initializes the Q-network, includes agent-environment interaction, stores experiences, and iteratively updates the Q-network until convergence [64]. The way DQN combines Q-learning, experience replay, target Q-networks, epsilon-greedy exploration, and deep neural networks (DNNs) makes it possible for reinforcement learning to work well and stay stable in complicated settings[65].

4| Performance Parameters

Performance parameters are essential in diabetic foot (DF) detection systems to assess the model’s diagnostic capability. These metrics ensure that automated systems can accurately distinguish between at-risk patients and healthy individuals, thus improving clinical outcomes [66].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

$$\text{AUC-ROC} = \int_0^1 \text{TPR}(\text{FPR})\, \text{dFPR} \tag{6}$$

Where:

- TP: True Positives (accurately identified DF cases).
- TN: True Negatives (accurately identified cases of non-DF).
- FP: False Positives (misidentified DF cases).
- FN: False Negatives (missed DF cases) [67].

Accuracy measures the proportion of true results (both true positives and true negatives) among the total number of cases analyzed. Sensitivity, or recall, measures the ratio of true positives accurately detected by the model. Specificity determines how accurately the model identifies actual negatives. Precision measures the proportion of positive identifications that are actually correct. The AUC-ROC measures the ability of the model to distinguish between classes. A higher AUC indicates better performance [68].

These parameters are crucial in the medical field, particularly in diabetic foot diagnosis, where missing a true case (false negative) could lead to serious complications such as infection or amputation. Therefore, sensitivity is often prioritized to ensure that all potential positive cases are detected. However, high sensitivity without sufficient specificity may lead to false alarms, increasing patient stress and unnecessary interventions.

While accuracy is a commonly reported metric, it can be misleading when the dataset is imbalanced—a common issue in medical datasets where the number of healthy cases may greatly outnumber the diseased ones. In such cases, the F1-score, which balances precision and recall, provides a more meaningful evaluation. Similarly, AUC-ROC is particularly useful for comparing models across different thresholds, making it an excellent tool for understanding overall classification performance.

In summary, no single parameter is sufficient on its own. A combination of these indicators offers a more comprehensive and reliable assessment of model performance, ensuring clinical decisions are both accurate and safe.

Among the reviewed studies, F1-score and AUC are often prioritized due to the imbalanced nature of most medical datasets, especially in DF diagnosis. Although accuracy is widely reported, it is not sufficient alone. Sensitivity is crucial in minimizing the risk of missing critical DF cases, while specificity ensures accurate exclusion of non-cases. A combined evaluation across these metrics yields a comprehensive understanding of the model's diagnostic power.

5 | Diabetic Foot Relevant Research

This review encompasses studies published between 2018 and 2025, retrieved from major scientific databases such as IEEE Xplore, PubMed, and Google Scholar. The selected literature primarily focuses on the application of thermographic imaging and AI techniques, particularly ML and DL, for the detection and classification of DF complications.

Over the past few years, considerable research has been devoted to developing automated DF detection systems using AI-driven medical image analysis. This section critically examines the most influential contributions within this domain, categorizing them into two primary methodological approaches: shallow learning and deep learning.

Each study is assessed based on its objectives, data sources, sample characteristics, algorithms used, performance metrics, and clinical relevance. The goal is to trace the evolution of diagnostic models, highlight performance trends, and identify current limitations to guide future research efforts.

The section is structured as follows: Subsection 5.1 presents studies based on shallow learning models such as SVM, DT, and Random Forests (RF), while Subsection 5.2 explores advanced DL architectures including CNNs, RNNs, GANs, and Transformer-based models.

5.1| Shallow learning Techniques

There have been a few studies in recent years on the use of ML for the detection and prognosis of diabetic foot ulcers (DFUs). F. Khan et al. [69] used a new ML approach that integrated reinforcement learning to enhance the analysis of DFU images, obtaining a classification accuracy of 92.5% and a remarkable efficiency gain of 78.45% in comparison to conventional methods. Sh. Hong et al. [70] concentrated on predicting DFU recurrence using the patient's risk factors; the highest accuracy rate obtained in the study was (93%) reported by the SVM model. However, the study suffers from a lack of dataset size and small feature variety in terms of features.

M. Alzyoud et al. [71] considered several classifiers and feature selection techniques on two datasets (DRD [72] and HCUP [73]), and indicated that the performance of the classifiers depend on the type of data used (Bayes Net [74], NB, Star [75], Multi Class Classifier [76], RF, Simple Logistic [77]). S. Stefanopoulos et al. [78] used CTREE and random forest models in order to predict the risk of major amputation in DFU, emphasizing gangrene and systemic infection as significant predictors. N. Arteaga-Marrero et al. [80] trained SVM classifiers based on thermographic data from the STANDUP [81], INAOE [82], and local [83] datasets, showing F1-scores of up to 90.27%. V. Filipe et al. [84] proposed two ML models to classify thermograms among diabetic severity stages, where Model 2 attained an accuracy of 93.2% . R. Alfkey et al. [85] used CNN (VGG-19) and PCA for feature representation and VGG-19 CNN [86], gradient boosting classifier [87], XGBoost [88], and RF for classification and achieved accuracy greater than 94%. Similarly, S. Kumar et al. [89] used several ML algorithms to test PIMA dataset [90] where logistic regression turned out with 80% accuracy.

Other studies have investigated image-based detection approaches. A. R. Naidu et al. [91] used MATLAB [92] for thermal image analysis, which allows non-contact acquisition of DF-related variations. J. Guzaitis et al. [93] published a smartphone-based screening model [94] that employed thermal imaging and edge detection to achieve more than 94% accuracy in the detection of inflammation. To assess the performance of an ML-based scoring scheme in identifying DF A. Khandakar et al. [95] applied the Synthetic Minority Oversampling Technique (SMOTE) [96]. CNNs and ML classifiers were benchmarked with dual-foot thermograms and image enhancement methods, and MobileNetV2 and DenseNet201 achieved the best balanced performances. Lastly, J. Saminathan et al. [97] experimented with thermal and color images from a FLIR E50 thermal imaging camera [98]. They developed a textured-temperature-based algorithm for the early detection of ulcer-prone regions with a detection accuracy of 95.61%.

Table 3. Summary of key studies investigating diabetic DF detection using various shallow machine learning ML techniques. The table outlines essential aspects, including the study aims, dataset types and sources, number of patients and images, methods employed, reported remarks.

chniques used to detect DF.

Ref	Aims	Data set: Type	Data set: Source	Data set: Number of patients	Data set: Number of images	Method	Remarks
[69]	Classify DFU using practical ML with reinforcement learning	DFU-Image Data set	Not Reported	Not Reported	Not Reported	Reinforcement Learning, ML	Achieved up to 78.45% improvement over traditional methods; suitable for remote, cost-effective, and real-time DFU screening. Dataset details and real-world validation were not provided.
[70]	Predict DFU recurrence in elderly diabetics	Clinical patient data	Hospital	138	Not Reported	SVM	The study's weaknesses include insufficient data collection, particularly for hereditary characteristics, which may impede the model's capacity to consider all variables affecting the recurrence of DFUs.
[71]	Compare ML classifiers on diabetic datasets	DRD and HCUP datasets	Diabetic Retinopathy Debrecen (DRD)[72], HCUP [73]	DRD: 1151, HCUP: 13067	1151 (DRD)	Bayes Net [74], NB, Star [75], MultiClass Classifier [76], random forest (RF), and Simple Logistic [77]	They used WEKA software to run the classifiers with default settings. They applied 10-fold cross-validation to validate the model's performance. The study was limited by using only traditional ML methods, without exploring DL models, which could have improved prediction accuracy, especially for large datasets such as HCUP. In addition, the dataset balance between the number of cases and features needs to be improved to ensure more robust predictions.
[78]	Predict major amputations in DFU patients	Inpatient clinical data	National Inpatient Sample(NIS 2008–2014) [79]	326,853	Not Reported	CTREE, RF	Despite their effectiveness, the CTREE and RF models can be computationally demanding, which could restrict their use in clinical real-time applications.
[80]	Classify DFUs using thermal imaging features	Thermal images	STANDUP [81], INAOE [82], Local datasets [83]	INOAE: 167, STANDUP: 227, and local: 22	Not Reported	SVM, RF, DL	Depending on the dataset and methodology used, the SVM classifier demonstrated significant variation in distinguishing diabetic and healthy individuals, especially when using advanced features from thermal images.

Table 3. (continued)

Ref	Aims	Data set				Method	Remarks
		Type	Source	Number of patients	Number of images		
[84]	Classify DF severity using thermograms	Thermal foot images	Public dataset of foot planter thermograms	120	Not Reported	KNN, SVM, LR	The proposed models effectively differentiate between healthy and diabetic foot thermographs, potentially aiding medical professionals in identifying injury risks and preventing foot ulcers.
[85]	Cluster and classify DFU severity using thermograms and ML	Thermogram images	Labeled diabetic thermogram dataset	Not Reported	Not Reported	VGG-19 [86], PCA, gradient boost classifier [87], the XGBoost classifier [88], and the RF, KNN	The authors used simple thermogram images, which didn't require a complex network for information extraction. They classified the images into severity classes based on the TCI score and medical professionals' judgment, potentially introducing subjectivity and inconsistencies in the classification process.
[89]	Evaluate ML models for diabetes onset prediction	Structured clinical data	PIMA Indian dataset [90]	768 (268 diabetic, 500 non-diabetic)	Not Reported	LR, ANN, KNN, DT, SVM, RF	The LR model demonstrated superior effectiveness in classifying and ensemble compared to the Zero-R model, although its small dataset may limit its generalizability and robustness.
[91]	Real-time thermal analysis of diabetic foot tissue changes	Thermal imaging	Cryogenic thermal camera data	Not specified	Not Reported	MATLAB (morphological processing, thresholding) [92]	The study lacks details on sample size, selection criteria, and ethical approval. Additionally, the approach is not compared to existing benchmarks or techniques for thermal imaging-based diabetes detection.
[93]	Detect foot outlines and inflammation via mobile thermal imaging	Thermal images	iPad Air + FLIR OnePro [94]	Not real patients (ghost feet used)	168	Edge detection, temperature thresholding	The method, using a symmetry analysis algorithm, achieved 94.28% accuracy in identifying inflammatory zones and 95.83% accuracy foot outlines and detecting potential health issues through temperature comparisons.

Table 3. (continued)

Ref	Aims	Data set				Method	Remarks
		Type	**Source**	**Number of patients**	**Number of images**		
[95]	Evaluate scoring and CNN models for DFU detection	Thermogram images + clinical data	122 diabetics, 45 healthy subjects	167	167 foot thermograms	DenseNet201, MobileNetV2, SMOTE [96], XGBoost, RF, AB	The research's applicability may be limited as it did not assess the clinical utility or practicality of the models in real-time scenarios like telemedicine or smartphone applications.
[97]	Early detection of DF via asymmetry in texture and temperature	Thermal + color images	FLIR E50 camera [98]	Not Reported	Not Reported	KNN, SVM; texture + thermal asymmetry	The study suggests a promising method for enhancing detection accuracy by combining texture and temperature features, but its small dataset may limit its findings.

This subsection reviews key shallow learning models applied in DF detection. As shown in Table 3, studies employed techniques such SVM, DT, RF, and LR. Most studies relied on thermal imaging or structured clinical data, with datasets ranging in size and quality. While some achieved high accuracy (e.g., Alfkey et al. at 95.08%), many lacked real-world validation or suffered from small sample sizes. These limitations suggest the need for broader datasets and external validation for future applications.

Table 4 presents the performance evaluation metrics used across the reviewed studies. It highlights the diagnostic strength of each approach in terms of accuracy, sensitivity, specificity, precision, recall, F1-score, and AUC, providing a comparative overview of model effectiveness in detecting diabetic foot complications.

Table 4. Performance evaluation of ML approaches of detecting diabetic foot.

Ref	Accuracy	Sensitivity	Specificity	Precision	Recall	F1-score	ROC-AUC
[69]	92.5%	71–98.2%	N/A	N/A	N/A	N/A	N/A
[70]	93%	92%	N/A	92%	92%	N/A	N/A
[71]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
[78]	77.7%	76.1%	79.3%	N/A	N/A	N/A	N/A
[80]	N/A	N/A	N/A	N/A	N/A	90.27%	N/A
[84]	93.2%	86.9%	95.4%	N/A	N/A	86.7%	N/A
[85]	95.08%	95.08%	95.09%	97.2%	N/A	95.08%	N/A
[89]	80%	N/A	N/A	N/A	N/A	N/A	68%
[91]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
[93]	94.28%	N/A	N/A	N/A	N/A	N/A	N/A
[95]	N/A	95.71%	N/A	N/A	N/A	N/A	N/A
[97]	95.61%	N/A	N/A	N/A	N/A	N/A	N/A

By organizing the findings in tabular format, this section enables a direct comparison of Ml techniques, datasets, and performance outcomes, providing a foundation for identifying future research directions.

In summary, the reviewed shallow learning studies show promising results in DF detection, especially when using well-known classifiers like SVM and Random Forest. However, limitations such as small sample sizes, lack of external validation, and computational complexity persist. Future research should focus on integrating these models into real-time clinical workflows and expanding datasets for better generalization.

5.2| DL Techniques

Current studies includes various DL-based methods for DFU detection, diagnosis, and risk assessment. P. L. Li et al. [99] introduced a novel deep learning framework based on DiffusionNet with integrated self-attention and anatomical features, with an accuracy of 82.9%, that outperforms the baseline models and used on a 3D foot scan dataset acquired from an EinScan Pro HD [100]. Similarly, V. Panamint et al. [102] used ResNet50V2 with plantar thermography, obtaining an accuracy of 71.8% and the potential of screening, with low amount of data and the integration of clinical variables.

L. Z. Chee et al. employed wearable sensor data. [103] inserted an approach combining CNN and LSTM models, which was 91:25% accurate but depended on the acceleration of walking. A. M. El-Kady et al. improved DFU detection with application of ResNet50 in pair with GANs of 84% diagnostic accuracy but lacked transparency of used datasets [104].

Gulshan and Arora [105] employed thermal imaging and a CNN-RNN pipeline, achieving 97.14% accuracy but without comparative baselines and validation. However, M. H. Alshayeji et al. [106] employed classic feature extractions (SIFT [107], SURF [108], BOF [109]) with SVM classification, achieving an accuracy of 97.81% but missed contemporary DL lifting.

R. N. Yousef et al. [110] presented a CNN-FS technique for the thermal image analysis with the accuracy of 99.3%, J. Reyes-Luévano et al. [111] introduced a new approach, DFU VIRNet, using estimated maps [112] trained on multimodal data, with AUC scores of over 0.99 for DFU and ischemia detection. N. Sharma et al. [113] used thermal-visual HSV image fusion blended with mask-RCNN segmentation to get the 92.5% agreement with the clinical evaluations.

Advanced models like vision transformers were also studied by H. Shao [114] achieved an accuracy of up to 99%, although the evaluation benchmarks were insufficient. Sh. Sh. Reddy et al. [115] performed a comparison of DL architectures and VGG16 managed an accuracy of 99.51% but with little information on its architecture. Similarly, M. Ahsan et al. [116] tested multiple CNNs on DFU2020 and found that ResNet50 achieved the highest precision in ischemia.

A. Hernandez-Guedes et al. [117] used variational dropout and SMOTE-boosted thermograms with 90% F1-score. P. N. Thotad et al. [118].They employed EfficientNet [119] and obtained an accuracy of 98.97% for DFU detection. V. Khullar et al. [120] reported better performance of the Inception ResNet V2 compared to conventional ML-based methods to classify DF from thermograms.

I. Khosa et al. [121], who achieved greater accuracy through custom CNN models on patch-level and full thermograms. M. Ray et al. [122] used asymmetric analysis [123] and Inception ResNet V2 on thermal images to drive mobile dependent DFU detection. Kh. Munadi et al. [124] presented a lightweight MobileNetV2–ShuffleNet fusion architecture, which attained 100% with classification rate.

Sh. Muralidhara et al. [125], that takes advantage of deep learning models for ulcer classification with a multi-class differentiation task improving significantly the ulcer grading by obtaining an accuracy of 0.9827. A. Anaya-Isaza et al. [126] suggested a temperature classification index (TCI) to categorize subjects based on temperature variation according to the following equation: (7) and more sophisticated augmentation techniques: 100% of detection.

$$\mathrm{TCI} = \frac{\mathrm{T} \cdot \mathrm{A}^2}{\mathrm{TCIR} + \mathrm{TCIL}} \tag{7}$$

where:

- T is the average temperature of both feet.
- A is the subject's age.
- The TCIR and TCIL are the thermal change indices for the right and left feet, respectively.

H. Maldonado et al. [127], has two cameras — Venice extension system (VS) camera [129] and IR thermo-camera [130] — on both feet forming the dual focus shot, and processes the acquired image automatically to detect necroses or sores. They combined Mask R-CNN VD[128] with dual-view thermal imaging for a mobile DFU detection system[131]. S. Assistant et al. [132] improved image pre-processing and thermal analysis through clustering [133], threshold [134], compression-based [135], and watershed transformation [136] with CNNs but without specific results. A. Bougrine et al. [137] used the U-Net [138] for thermal image segmentation and risk assessment which was demonstrated during a prospective clinical study on 122 patients in de Mayo National Hospital [139]. L. Alzubaidi et al. [140], where DFU_QUTNet was proposed using DAG model [141]. DFU_QUTNet was composed of input layers with three channels. All channels were of size 224 × 224 pixels. The processing of the convolutional layer through the previous layer involved convolution with a learnable filter set, batch normalization (BN) [142], rectified linear unit (ReLU) [143], addition, average pooling, dropout, and fully connected (FC) layers [144]. The features were extracted by the DFU_QUTNet network to train the SVM and KNN classifiers. For comparison, it reached a 94.5% f1-score, which was higher than the other CNNs.

M. Goyal et al. [145] used EfficientDet on augmented DFU images and achieved high accuracy without performance numbers. M. Kayalvizhi et al. [146] also suggested CNN+SVM hybrid models which obtained 97.9%Classification Accuracy. Finally, Cruz-Vega et al. [147] presented a DFTNet model that employed Fourier transform-based feature extraction and achieved better performance than AlexNet or GoogleNet, with a sensitivity of up to 95.34%; however, the sample size was restricted in terms of a multi-level classifier. M. Goyal et al. [148] also reported DFUNet as a CNN for accurate DFU classification (AUC = 0.961) with promising potential for automated diagnosis and clinical use.

DL techniques, particularly CNNs, have shown remarkable performance in the automatic detection of DF complications. Despite this, several challenges persist, including limited annotated data, high computational requirements, and a lack of model transparency. Emerging studies highlight the benefit of combining DL architectures with preprocessing methods and ensemble learning strategies to improve diagnostic accuracy and robustness.

Table 5. Comparative summary of key studies employing DL techniques for DF detection. The table outlines the study aims, dataset characteristics, DL models applied, and reported limitations, providing insights into current approaches and challenges in DF diagnostic automation.

In summary, DL models have achieved high accuracy across various datasets and imaging modalities. However, the generalizability of these models remains constrained by dataset imbalance, lack of clinical context, and interpretability issues. While architectures such as ResNet, VGG, and EfficientNet dominate the field, hybrid approaches and attention-based models are emerging as powerful alternatives. There remains a clear need for standardized datasets, longitudinal clinical validation, and explainable DL systems to support clinical adoption.

Table 6. Overview of performance evaluation metrics reported in DL-based studies for diabetic foot detection. Metrics include accuracy, sensitivity, specificity, precision, F1-score, and ROC-AUC, enabling comparison of diagnostic effectiveness across different models.

Table 5. Comparative summary of key studies employing DL approaches for DF detection.

Ref	Aims	Data set				Method	Remarks
		Type	Source	Number of patients	Number of images		
[99]	Classify six diabetic foot types using 3D scanned foot images enhanced with DiffusionNet and Self-Attention.	3D Scanned Images + External Measurements	EinScan Pro HD Scanner [100]	114 diabetic patients (50 male – 64 female)	Not Reported	DiffusionNet + Self-Attention, SVM, RF, XGBoost, LightGBM, DGCNN, PointNet++ [101]	Achieved 82.9% accuracy with strong F1-score and low classification error (4–10 misclassifications over 5-fold CV). Sample size is relatively small; not yet validated in real-world applications. Architecture improvements are still needed.
[102]	To develop a deep learning model based on thermographic foot images to classify diabetic patients by risk level (IWGDF).	Plantar Thermal Images	tCam-mini thermal camera (320×240), Ramathibodi Hospital, Thailand	Adults with diabetes; total 153 images	153 (Train: 98, Test: 55)	ResNet50V2 (transfer learning), OpenCV preprocessing, SPSS for statistical analysis	A relatively small data set, the absence of longitudinal outcome tracking, and the exclusion of several clinically relevant factors. Future research should incorporate a more diverse and representative dataset, integrate additional predictive variables, and enhance model interpretability to support clinical decision-making and facilitate real-world adoption.
[103]	Detect diabetes using CNN-LSTM from wearable sensor acceleration data.	Walking acceleration data	Wearable sensors on hip, knees, ankles	Not Reported	Not Reported	CNN-LSTM vs CNN, LSTM	The technique uses walking acceleration data, potentially omitting relevant artifacts. Additionally, the use of an anonymized dataset could lead to concerns about the validity of the results.

Table 5. (continued).

Ref	Aims	Data set				Method	Remarks
		Type	Source	Number of patients	Number of images		
[104]	Improve DFU identification using ResNet50 and GANs on DFU images.	Colored DFU images	500 images (unspecified dataset)	Not Reported	500	ResNet50; ResNet50+GAN	The paper offers a hybrid model for improved diagnosis of DFU in clinical settings, demonstrating significant progress in DFU management and treatment.
[105]	Non-invasive DM detection using RNN on thermal foot images.	Thermal foot images	50 volunteers (diabetic and non-diabetic)	50	Not Reported	CNN for features, RNN for prediction	It is more difficult to assess the efficacy and superiority of the authors' RNN methodology due to the absence of comprehensive evaluation metrics and comparisons with other approaches.
[106]	Use BOF [107]+SIFT [108]+SURF [109] with thermal images for early DFU diagnosis.	Thermal plantar images	Plantar thermogram DB	Not Reported	Not Reported	SURF-BOF + SVM	The proposed method uses conventional ML techniques, potentially less effective than DL methods. Also, the manually created features from SIFT and SURF, are potentially unsuitable for different image types or data variations.
[110]	Develop CAD system using deep CNN on thermal images.	Thermal foot images	Not Reported	Not Reported	Not Reported	Deep CNN	The study's small and imbalanced dataset, with 10 participants per class, may potentially bias the model towards the majority and affect the accuracy of less-represented classes.
[111]	Introduce DFU_VIRNet [112] to classify DFU, ischemia, infection using multiple datasets.	Visible & thermal images	Five datasets incl. IR	Not Reported	Not Reported	DFU_VIRNet (CNN + GAP-2D-DLSA-IMG)	The DFU_VIRNet mechanism enhances model cognitive abilities but does not provide a method for identifying and visualizing DFU-risk zones based on deep CNN features for classification.

Table 5. (continued).

Ref	Aims	Data set: Type	Data set: Source	Data set: Number of patients	Data set: Number of images	Method	Remarks
[113]	Assess HSV-based fusion with Mask-RCNN for wound segmentation.	Thermal + visual images	42 patients, 12-week study	42	Not Reported	HSV fusion + Mask-RCNN	The study's small dataset may limit the generalizability of its results.
[114]	Improve image classification using enhanced Vision Transformer with a learnable block.	Medical image datasets	Not Reported	Not Reported	Not Reported	Enhanced Vision Transformer	The method demonstrated the potential utility of vision transformers for handling small-size medical image datasets.
[115]	Early detection of foot ulcers using DL models including VGG16.	Foot ulcer images	1029 ulcer-specific images	Not Reported	1029	ResNet, VGG16, DenseNet, MobileNet	The study focused on the practical application of algorithms for predicting DFU and identifying stages, contributing to medical diagnostics and prognostics.
[116]	Classify infection and ischemia using CNN-based DL models on DFU2020 dataset.	DFU2020 benchmark dataset	Public dataset	Not Reported	Not Reported	AlexNet, VGG16/19, GoogLeNet, ResNet50/101, MobileNet, SqueezeNet, DenseNet	The research proposes affine transformation methods for data augmentation, which may enhance model performance; however, it does not provide qualitative analysis or visualization of learned features, which could aid in comprehending the underlying mechanisms and elucidating model predictions.
[117]	Improve DFU classification using thermogram data with SMOTE and variational DL.	Thermogram images	INAOE dataset + private thermograms + SMOTE synthetic	Not Reported	Not Reported	Variational DL + LASSO + RF + SVM	The proposed feature selection methods by the authors may be constrained by the utilization of a small, unbalanced dataset of thermograms from diabetic and healthy individuals.
[118]	Early DFU detection using EfficientNet [119] hybrid deep CNN model.	Foot images (normal and diabetic)	844 images	Not Reported	844	EfficientNet vs AlexNet, GoogleNet, VGG16/19	The level of complications or the intensity of discomfort cannot be determined online.

Table 5. (continued).

Ref	Aims	Data set				Method	Remarks
		Type	**Source**	**Number of patients**	**Number of images**		
[120]	Compare DL vs ML methods in DF classification using thermograms.	Plantar thermograms	Not Reported	Not Reported	Not Reported	DL: CNNs, Inception ResNet V2; ML: SVM, RF, etc.	The study highlighted the significance of data augmentation enhancing the effectiveness of the models employed for diabetic foot condition classification.
[121]	Use CNN models with thermograms at image and patch levels for DFU detection.	Plantar thermogram images	334 thermograms (control + diabetic)	Not Reported	334	ML & DL (ResNet50, DenseNet121, custom CNN)	The CNN-based model outperformed other models and techniques, and using image-level thermogram data increased recognition accuracy for both ML and DL methodologies.
[122]	Non-invasive DFU detection using infrared imaging and asymmetric analysis [123].	Thermal foot images	60 participants, True IR U5855A camera	60+ (average age 50–60)	Not Reported	CNN, Inception ResNet V2, MATLAB image analysis	The study relied on dividing the foot into six parts, and 'asymmetric analysis' was used to track the temperature change. It does not specify the exact algorithm used for registration.
[124]	DFU classification using decision fusion with MobileNetV2 and ShuffleNet.	Plantar thermogram images	Not Reported	Not Reported	Not Reported	MobileNetV2, ShuffleNet, Decision Fusion	The authors utilized image augmentation to balance the dataset, suggesting that larger, diverse data sets could improve the generalization and performance of DL models.
[125]	CNN for multi-class classification and severity grading of DF.	Plantar thermal images	Not Reported	Not Reported	Not Reported	Custom CNN vs AlexNet	The paper presents a method for predicting ulceration sites using thermogram data; however, it is deficient in qualitative analysis and visual explanations for interpreting thermograms and identifying ulceration spots.
[126]	Foot thermography + DL with advanced augmentations for DM detection.	Foot thermography images	TCI-based classification, various augmentation methods	Not Reported	Not Reported	DL, PCA, ICA, Gaussian filter, LDA, dictionary learning	The results' generalizability and validity may have been compromised due to the lack of comparison between proposed data augmentation techniques and existing practices or clinical diagnoses.

Table 5. (continued).

Ref	Aims	Data set: Type	Data set: Source	Data set: Number of patients	Data set: Number of images	Method	Remarks
[127]	Thermal image segmentation with Mask-RCNN [128] and mobile app development.	Foot thermal images	141 segmented images, Venice system (VS) + IRT cameras [129-130]	Not Reported	141	Mask-RCNN, dual-camera thermal capture, mobile implementation [131]	To achieve the best results, a comprehensive dataset composed of network images must be used to develop a high-quality classification system. Issues with data quality and standardization can impact their performance.
[132]	Thermal DF detection using DL and image segmentation techniques.	Thermal foot images	Not Reported	Not Reported	Not Reported	CNN + preprocessing and segmentation (clustering, watershed, threshold, and compression-based) methods [133-136]	The authors present a sophisticated technique for processing thermal images of diabetic feet, promising for healthcare applications, potentially improving the detection and management of DFUs.
[137]	Segment plantar thermal images with U-Net [138] and classify foot ulcer risk.	Thermal + color foot images	FLIR ONE Pro; clinical study in Peru's Dos de Mayo National Hospital [139]	122 T2D patients	Not Reported	U-Net segmentation; risk classification (R0, R1, R2)	The authors proposed a method to segment plantar foot regions from thermal images and analyze temperature distribution, but this may not be effective with diverse foot sizes, shapes, orientations, or backgrounds not represented in the training data.
[140]	Propose DFU_QUTNet CNN to classify DFU vs normal skin using DAG (Directed Acyclic Graph) design [141].	Foot ulcer skin images	754 images (DFU + healthy)	Not Reported	754	DFU_QUTNet, SVM, KNN; compared to GoogleNet, VGG16, AlexNet [142-144]	The paper lacks qualitatively analyzed or visualized features learned by DFU_QUTNet, which could help understand how the network differentiates between diseased and normal skin.
[145]	Use EfficientDet DL model on 4500 DFU images for improved detection.	DFU images	4500 images; DFUC dataset	Not Reported	4500	EfficientDet + preprocessing + augmentation	The EfficientDet algorithm was found effective in detecting diabetes, but no evaluation metrics were provided beyond MAP.

Table 5. (continued).

Ref	Aims	Data set				Method	Remarks
		Type	Source	Number of patients	Number of images		
[146]	Detect diabetes using CNN+SVM hybrid model on foot thermal images.	Thermal foot images	Split into train/test sets	Not Reported	Not Reported	CVM (CNN+SVM hybrid), cross-validation	Hybrid DL algorithms for diabetes prediction using thermal imaging show potential for improving patient outcomes but may overfit training sets and underperform in unusual circumstances.
[147]	Propose DFTNet for thermal image classification using DFT + DL.	Thermal plantar images	Not Reported	Not Reported	Not Reported	DFTNet, ANN, SVM, graph-based ROI, AlexNet, GoogleNet	The proposed system, which divided images into five levels based on TCI values, reduced the number of samples per class, which made it unsuitable for classifiers.
[148]	DFUNet CNN for DFU vs healthy skin detection in foot images.	DFU skin images	Custom DFU dataset	Not Reported	Not Reported	DFUNet, CNNs, traditional ML	Healthcare practitioners can utilize the DFUNet model as a diagnostic tool to assist them in treating DFUs in a timely and accurate manner.

Table 6. Performance evaluation of DL approaches of detecting diabetic foot.

Ref	Accuracy	Sensitivity	Specificity	Precision	F1-score	ROC-AUC
[99]	82.9%	N/A	N/A	N/A	N/A	N/A
[102]	71.8%	81.2%	64.0%	N/A	N/A	N/A
[103]	91.25%	N/A	N/A	N/A	N/A	N/A
[104]	84%	N/A	N/A	85%	0.84	N/A
[105]	97.14%	N/A	N/A	N/A	N/A	N/A
[106]	N/A	N/A	N/A	97.9%	N/A	0.9995
. [110]	99.3%	N/A	N/A	N/A	N/A	N/A
[111]	N/A	N/A	N/A	N/A	0.9600	0.9923
[113]	92.5%	N/A	N/A	N/A	N/A	N/A
[114]	99%	N/A	N/A	N/A	N/A	N/A
[115]	99.51%	N/A	N/A	N/A	N/A	N/A
[116]	N/A	N/A	N/A	99.49%	N/A	N/A
[117]	N/A	N/A	N/A	N/A	0.90	N/A
[118]	99%	N/A	N/A	N/A	98%	N/A
[120]	N/A	N/A	N/A	N/A	N/A	N/A
[121]	N/A	N/A	N/A	N/A	N/A	N/A
[122]	N/A	N/A	N/A	N/A	N/A	N/A
[124]	100%	N/A	N/A	N/A	N/A	N/A
[125]	98.27%	96.84%	98.92%	N/A	N/A	N/A
[126]	100%	N/A	N/A	N/A	N/A	N/A
[127]	90% (ulcers)	N/A	N/A	N/A	N/A	N/A
[132]	N/A	N/A	N/A	N/A	N/A	N/A
[137]	N/A	N/A	N/A	N/A	N/A	N/A
[140]	N/A	N/A	N/A	N/A	0.945	N/A
[145]	N/A	N/A	N/A	N/A	N/A	N/A
[146]	97.9%	93.6%	N/A	N/A	N/A	N/A
[147]	94.53%	95.34%	93.75%	N/A	N/A	N/A
[148]	N/A	N/A	N/A	N/A	N/A	0.961

6| The Main New Directions of Research in The Detection of DF

6.1| Directions In Image Acquisition

X-rays have traditionally been used to detect DF, but other imaging systems have also been used. For example, infrared columns, ultrasound, magnetic resonance imaging, anatomy, and thermography have been used to obtain more accurate information.

Recent studies are increasingly exploring infrared technology for medical imaging due to its non-invasive nature and absence of radiation, which enhances safety in diagnosis. Moreover, it is more accurate and can detect subtle temperature changes that may indicate the presence of a disease. Additionally, the combination of thermography, a non-invasive technique, and advanced molecular analysis has shown promising results in the early detection of diabetes. More research is expected to take place using infrared thermal images.

6.2| Future Directions in ML and DL

This research compilation from 2018 to 2025 examined diverse methodologies for diagnosing and classifying DFUs through thermal imaging techniques. For the accurate and timely detection of DFUs, the reviewed studies have used CNNs, hybrid models, and traditional ML models, which led to significant advancements. These methods concentrate on augmenting accuracy in DFU detection through the optimization of model architecture, the advancement of algorithms, and the enhancement of dataset quality. DL models, including EfficientNet and DFU_QUTNet, and hybrid approaches such as CNN-LSTM demonstrate significant efficacy in the analysis of small medical image datasets.

Using methods like data augmentation and transfer learning, along with adding local thermograms and creating synthetic data, has improved model performance by making up for the lack of data. Furthermore, multi-level image analysis that emphasizes both image-level and patch-level thermogram data has arisen as an effective method for identifying DFUs. The integration of ML with DL through hybrid models like CNN-SVM has improved accuracy in early detection, leveraging DL's feature extraction and ML's decision-making capabilities. Moreover, exploring variational DL techniques, including variational dropout, has been acknowledged as an effective strategy for improving model generalization and robustness in scenarios with constrained datasets. Lastly, the diversity and quality of datasets remain pivotal, with studies emphasizing the need for balanced datasets to avoid biases and enhance model performance across diverse populations.

6.3| Directions in the Dataset

We recommend that researchers employ recognized benchmark datasets, such as those available in TREC [149], in IR research (information retrieval), emphasizing the significance of embracing a shared and widely acknowledged framework for evaluating algorithms. This approach fosters transparency, facilitates fair comparisons, and drives advancements in the field through collaborative endeavors. Encouraging researchers to adhere to these standards is crucial for making meaningful contributions to the broader scientific community.

7| Clinical Implications and Deployment Challenges

However, despite the significant potential offered by recently developed ML and DL models for the automatic detection and categorization of DF complications, their practical clinical implementation is scarce. Although many studies have reported high diagnostic performance in terms of accuracy, sensitivity, and specificity, few have been validated in real-world clinical settings, including hospitals and telemedicine sites. This raises serious concerns about the generalizability, stability, and clinical readiness of such models in different patient populations and hospital contexts.

There are several bottlenecks to deploying AI-based diagnostic systems in clinical practice.

- Variability of Data: Most models are developed from limited or homogeneous datasets and cannot fully capture age differences, skin tone variations, ulcer stages, and/or comorbidity. This limits the generalizability of the model to a larger patient pool.
- Workflow Integration: For AI tools to be successfully integrated into clinical workflows, they must integrate seamlessly with HISs, possess user-friendly interfaces for medical personnel, and comply with data privacy laws and regulations.
- Model Interpretability and Clinical Trust: Most DL models are “black boxes” and provide little insight into the reasoning behind decisions. However, it lacks visual explanations or interpretable outputs (e.g., heatmaps), which may lower clinician trust and limit its applications.
- Resource constraints: The computational requirements of DL models can be a bottleneck in environments where resources are scarce, such as rural or underprivileged areas. One such challenge is the development of lightweight, mobile-friendly models.

To overcome these pitfalls, future research should focus on the following strategies:

- Conducting multicenter clinical trials to validate model performance in various healthcare environments and patient populations.
- Developing models that are efficient and lightweight to perform in real time on mobile or embedded devices to increase accessibility in low-resource settings.

- Promoting interdisciplinary research: AI researchers and physicians might work together to explore whether model outputs match clinicians' requirements and clinical jargon.
- In addition to algorithmic performance, successful clinical adoption of AI-powered DF detection systems relies on practicality, acceptance by clinicians, and patient safety. There is a need to narrow the gap between research and clinical practice to accomplish real progress in early detection, early intervention, and patient survival.

8| Conclusions

In this study, we provide an up-to-date review of the state-of-the-art techniques for diabetic foot (DF) detection using infrared thermal imaging and machine learning (ML) and deep learning (DL) methods. It systematically investigates 40+ studies regarding techniques, datasets, and criteria published from 2018 to 2025. Important challenges, such as data imbalance, absence of dataset standardization, small clinical validation, and applicability of the findings in a real-world scenario, are deeply analyzed. The conclusions of this study provide directions for further research to improve diagnostic accuracy in the transition from animal models to the realization of AI-based diabetic foot screening tools in the clinic.

Recent advances in ML and DL techniques have notably enhanced the performance of computer-aided diagnostic systems, particularly in medical image analysis. Traditional ML models, such as SVM and LR, have shown good performance when combined with feature selection methods, such as Gini impurity and information gain. In addition, CNN networks (ResNet, DenseNet, VGG16/19) and custom models (DFUNet, etc.) were also found to perform well in detecting DFUs from thermal images. Hybrid models that combine CNN with LSTM or SVM classifiers exploit both temporal and spatial features, achieving better results.

Furthermore, data augmentation methods, including SMOTE, and regularization approaches, such as L1 and L2, are utilized to combat overfitting and improve generalization, especially when applied to small and imbalanced datasets. Advanced DL-based methods, such as EfficientDet and Vision Transformers, are also gaining strength in challenging medical imaging tasks. Cross-validation is a fundamental step in model evaluation, as it contributes to increasing robustness and reproducibility. Overall, the amalgamation of these AI-driven tools has great promise in augmenting early DF detection, ultimately enhancing patient prognosis and reducing the economic burden on healthcare services.

Author Contribution

All authors contributed equally to this work.

Funding

This research has no funding source.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

References

[1] Ministry of Health & Family Welfare India, "The DF : Prevention and Management in India Ministry of Health & Family Welfare Government of India" Natl. Stand. Treat. Guidel., vol. 0, pp. 0–74, Aug 2017.

[2] M. R. Abouzid, K. Ali, I. Elkhawas, and Sh. M. Elshafei, "An Overview of Diabetes Mellitus in Egypt and the Significance of Integrating Preventive Cardiology in Diabetes Management" Cureus, vol. 14, no. 7, July 2022, doi: 10.7759/cureus.27066.

[3] I. W. Suryasa, M. Rodríguez-Gámez, and T. Koldori, " Health and Treatment of Diabetes Mellitus", International Journal of Health Sciences, vol. 5, no. 1, pp. i-v, April 2021, doi: https://doi.org/10.53730/ijhs.v5n1.2864.

[4] R. Kumar, P. Saha, Y. Kumar, S. Sahana, A. Dubey, and O. Prakash, "A Review on Diabetes Mellitus: Type1 & Type2" World Journal of Pharmacy and Pharmaceutical Sciences, vol. 9, no. 10, pp. 838-850, Sept 2020 doi: 10.20959/wjpps202010-17336.

[5] J. Gao, Y. Yang, P. Lin, and D. S. Park, "Editorial: Computer vision in healthcare applications" J. Healthc. Eng., vol. 2018, no. 5157020, pp. 1-4, 2018, doi: 10.1155/2018/5157020.

[6] A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, and R. Socher, "Deep learning-enabled medical computer vision", NPJ digital medicine, vol. 4, no. 1, pp. 1-5, Jan 2021, doi: 10.1038/s41746-020-00376-2

[7] American Diabetes Association, "Diagnosis and Classification of Diabetes Mellitus" Diabetes care, vol. 33, no. Supplement_1, pp. S62--S69, Jan 2010, doi: 10.2337/dc10-S062.

[8] K. Saarinen, "Image processing, analysis and machine vision" Springer, vol. 35, no. 1, pp. 102-104, 1994, doi: 10.1016/0165-1684(94)90202-x.

[9] L. K. Lee and S. C. Liew, "A study of medical image processing tools" in 2015 4th International Conference on Software Engineering and Computer Systems, ICSECS 2015: Virtuous Software Solutions for Big Data, pp. 171–176, 2015, doi: 10.1109/ICSECS.2015.7333105.

[10] M. Muthuraja and N. Shanthi, "A study on classification of DF ulcer using machine learning algorithms," Ann. Rom. Soc. Cell Biol, vol. 25, no. 1, pp. 1881–1894, 2021, doi: https://annalsofrscb.ro/index.php/journal/article/view/305.

[11] M. D. Grossberg and S. K. Nayar, "A general imaging model and a method for finding its parameters" Proc. IEEE International Conference Computing Vision., vol. 2, pp. 108–115, 2001, doi: 10.1109/iccv.2001.937611.

[12] https://www.nibib.nih.gov/science-education/science-topics/x-rays , last visited on Jun 2024.

[13] https://my.clevelandclinic.org/health/diagnostics/4995-ultrasound , last visited on 15/1/2024.

[14] https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/magnetic%20resonance-imaging-mri , last visited on 15/1/2024.

[15] N. Nölke, "Introduction to thermal imaging" Megger, vol. 44, no. 0, pp. 1–5, 2012.

[16] N. Dey, A. S. Ashour, and A. Althoupety, "Thermal Imaging in Medical Science" in Computer Vision: Concepts, Methodologies, Tools, and Applications, pp. 1-20, Jan 2017, doi: 10.4018/978-1-5225-5204-8.ch046.

[17] E. F. J. Ring and K. Ammer, "Infrared thermal imaging in medicine" Physiol. Meas., vol. 33, no. 3, pp. 1-15, 2012, doi: 10.1088/0967-3334/33/3/R33.

[18] K. J. Havens and E. J. Sharp, "Imager Selection" Therm. Imaging Tech. to Surv. Monit. Anim. Wild, pp. 121–141, 2016, doi: 10.1016/b978-0-12-803384-5.00008-7.

[19] R. Sensing, I. Spec-, K. J. Havens, and E. J. Sharp, "Imager Selection Using Thermal Imagers for Animal Ecology Thermal Imagers and System Considerations Contribution of Remote Sensing for Crop and Water Monitoring Properties of Thermal Signatures Thermal Imaging Applications and Experiments" Academic Press, pp. 121-141, 2016, doi: https://doi.org/10.1016/B978-0-12-803384-5.00008-7.

[20] D. S. Jayas and R. Vadivambal, "Thermal imaging" Food Eng. Ser., no. September, pp. 183–197, 2014, doi: 10.1007/978-1-4939-0311-5_8.

[21] L. Acampora, F. De Filippis, A. Martucci, and L. Sorgi, "3D Reconstruction of Thermal Images" 26th Aerosp. Test. Semin. 2011,Springer, no. May 2014, pp. 263–276, 2011.

[22] I. S. Anggreni, H. Kristianto, D. Handayani, Y. Yueniwati, P. L. T. Irawan, R. Rosandi, R. E. Kapti, and A. D. Purnama, "Artificial intelligence for diabetic foot screening based on digital image analysis: A systematic review," *Journal of Diabetes Science and Technology*, 2025. doi: 10.1177/19322968251317521.

[23] I. M. Ibrahim, and A. M. Abdulazeez, " The Role of Machine Learning Algorithms for Diagnosing Diseases", Journal of Applied Science and Technology Trends, vol. 2, no. 01, pp. 10-19, March 2021, doi: https://doi.org/10.38094/jastt20179.

[24] P. Cunningham, M. Cord, and S. Delany, "Supervised Learning" Springer, pp. 21–49, 2008, doi: 10.1007/978-3-540-75171-7_2.

[25] S. Siadati, "What is unsupervised Learning" Strategic Intelligence Research Lab (SIRLAB), Aug 2018, doi: 10.13140/RG.2.2.33325.10720.

[26] M. Devgan, G. Malik, and D. K. Sharma, "Semi-Supervised Learning" International Journal of Engineering & Technology, pp. 251-280, 2020, doi: 10.1002/9781119654834.ch10.

[27] M. Naeem, S. Rizvi, and A. Coronato, "A Gentle Introduction to Reinforcement Learning and its Application in Different Fields" IEEE Access, vol. 8, pp. 209320–209344, Jan. 2020, doi: 10.1109/ACCESS.2020.3038605.

[28] M. M. Ahsan, S. A. Luna, and Z. Siddique, "Machine-Learning-Based Disease Diagnosis : A" Healthcare, vol. 10, pp. 1–30, 2022, doi: 10.3390/healthcare10030541.

[29] L. Rokach and O. Maimon, "Decision Trees" Lect. Notes Math., vol. 1928, pp. 67–86, 2008, doi: 10.1007/978-3-540-75859-4_5.

[30] https://medium.com/@sachinsoni600517/mastering-decision-tree-models-for-smarter-decisions-451ef8b75836, last visited on 16/1/2024.

[31] H. Bittencourt, and R. Clarke, " Use of classification and regression trees (CART) to classify remotely-sensed digital images" 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings, vol. 6, pp. 3751-3753, 2003, doi: 10.1109/IGARSS.2003.1295258.

[32] https://scikit-learn.org/stable/modules/svm.html , last visited on 8/4/2023.

[33] A. Ng, "CS229 Lecture notes Margins : Intuition" Intell. Syst. their Appl. IEEE, vol. pt.1, no. x, pp. 1–25, 2000, doi: 10.1016/j.aca.2011.07.027.

[34] L. Ladichy, and P. Torr, " Linear Support Vector Machines" Proceedings of the 28th International Conference on Machine Learning, pp. 985-992, Jan 2011.

[35] M. Steinbach and P.N. Tan, "8-kNN" in machine learning algorithms, pp. 152–161, 2009.

[36] https://www.kdnuggets.com/2020/10/exploring-brute-force-nearest-neighbors algorithm.html, last visited on 10/4/2023.

[37] "L. Regression,18-661 Introduction to Machine Learning" ECE – Carnegie Mellon University, pp. 1-57, Spring 2020.

[38] C. R., P. Jeipratha, T. Daniya, and N. Kanagavalli, "Logistic Regression Model with Gradient Descent Parameter Estimation" Journal of Critical Reviews, vol. 7, no. 12, pp. 2048-2052, 2020, doi: 10.31838/jcr.07.12.306.

[39] D. Szafron, R. Greiner, P. Lu, D. Wishart, C. Macdonell, J. Anvik, B. Poulin, Z. Lu, R. Eisuer, and E. Ca, "Explaining naïve Bayes classifications", pp. 1-18, Jan 2003.

[40] S. Xu, Y. Li, and W. Zheng, " Bayesian Multinomial Naïve Bayes Classifier to Text Classification" International Conference on Future Information Technology, pp. 347-352, 2017, doi: 10.1007/978-981-10-5041-1_57.

[41] A. Khan, A. Sohail, U. Zahoora and A. S. Qureshi, "Review of Deep Learning Algorithms and Architectures," IEEE Access, vol. 7, pp. 53040-53065, 2019, doi: 10.1109/ACCESS.2019.2907012.

[42] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, O. Al-Shamma, Y. Duan, J. Santamaria, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions" Journal Big Data, vol. 8, no. 1, p. 74, 2021, doi: 10.1186/s40537-021-00444-8.

[43] A. Mathew, A. Arul, and S. Sivakumari, "Deep Learning Techniques: An Overview" Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020, pp. 599–608, 2021, doi: 10.1007/978-981-15-3383-9_54.

[44] R. Dastres and M. Soori, "Artificial Neural Network Systems" Int. J. Imaging Robot, International Journal of Imaging and Robotics (IJIR), vol. 21, pp. 13–25, Mar. 2021.

[45] M. G. Anderson D, "Artificial Neural Networks Technology", Kaman Sciences Corporation, vol. 258, no. 6, pp. 1–83, 1992.

[46] R. Grosse, "Lecture 5: Multilayer Perceptrons," pp. 1–7, 2018, [Online]. Available: https://www.cs.toronto.edu/~rgrosse/courses/csc311_f21/readings/Multilayer Perceptrons.pdf, last visited on 20/12/2023.

[47] L. Vanneschi and M. Castelli, "Multilayer perceptrons," vol. 1–3, pp. 153-261, 2018, doi: 10.1016/B978-0-12-809633-8.20339-7.

[48] F. M. Salem, F. Simple, and G. Architectures, "Recurrent Neural Networks", Springer International Publishing, pp. 1-119, 2022.

[49] K. Yao, G. Zweig, M. Y. Hwang, Y. Shi, and D. Yu, "Recurrent neural networks for language understanding," Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH, no. 1, pp. 2524–2528, 2013, doi: 10.21437/interspeech.2013-569.

[50] P. Bharati and A. Pramanik, "Deep Learning Techniques—R-CNN to Mask R-CNN: A Study," vol. 999, pp. 657-668, 2020, doi: 10.1007/978-981-13-9042-5_56.

[51] A. Patil and M. Rane, "Convolutional Neural Networks: An Overview and Its Applications in Pattern Recognition," Smart Innov. Syst. Technol., vol. 195, pp. 21–30, 2021, doi: 10.1007/978-981-15-7078-0_3.

[52] S. Srihari, "Convolutional Networks Overview," pp. 1–25, 2019, [Online]. Available: https://en.wikipedia.or, last visited on 12/1/2024.

[53] L. Lazebnik, "L01. Convolutional Neural Network Architectures: from LeNet to ResNet," CS 598 - Deep Learn. Recognit., 2016, [Online]. Available: https://slazebni.cs.illinois.edu/spring17/lec01_cnn_architectures.pdf, last visited on 13/12/2023.

[54] J. Brownlee, "A Gentle Introduction to Pooling Layers for Convolutional Neural Networks," Machine Learning Mastery, 2019, doi: 10.13140/RG.2.2.30828.10880

[55] https://www.analyticsvidhya.com/blog/2020/02/learn-image-classification-cnn-convolutional-neural-networks-3-datasets/, last visited on 19/12/2023.

[56] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, Sh. Ozair, A. Courville, Y. Bengio, " Generative Adversarial Nets" Advances in Neural Information Processing Systems 27 (NIPS 2014), vol. 44, pp. 9629-9640, 10 June 2014, doi: 10.1145/3422622.

[57] S. Wang, "Generative Adversarial Networks (GAN) A Gentle Introduction" 24 April 2018, doi: https://www.researchgate.net/publication/316382604, last visited on 12/1/2024.

[58] O. STRIUK and Y. KONDRATENKO, "Generative Adversarial Neural Networks and Deep Learning: Successful Cases and Advanced Approaches" International Journal of Computing, vol. 20, no. 3, 01 March 2022, doi: https://doi.org/10.47839/ijc.20.3.2278.

[59] H. E. Egilmez, A. K. Singh, M. Coban, M. Karczewicz, Y. Zhu, Y. Yang, A. Said, and T. S. Cohen, " Transform Network Architectures for Deep Learning based End-to-End Image/Video Coding in Subsampled Color Spaces" Electrical Engineering and Systems Science, pp. 1-10, 27 Feb 2021, doi: https://doi.org/10.48550/arXiv.2103.01760.

[60] https://en.wikipedia.org/wiki/Transformer_(machine_learning_model), last visited on 20/12/2023.

[61] Sh. Chen and W. Guo, "Auto-Encoders in Deep Learning—A Review with New Perspectives", vol. 11, no. 8, 2023, doi: https://doi.org/10.3390/math11081777.

[62] https://theprofessionalspoint.blogspot.com/2019/06/autoencoders-in-deep-learning.html , last visited on 20/12/2023.

[63] https://www.simplilearn.com/tutorials/deep-learning-tutorial/what-are-autoencoders-in-deep-learning , last visited on 20/12/2023.

[64] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, Ch. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, Sh. Legg and D. Hassabis, " Human-level control through deep reinforcement learning", Semantic Scholar, vol. 518, pp. 529-533, Feb 2015.

[65] M. Sewak, "Deep Q Network (DQN), Double DQN, and Dueling DQN: A Step Towards General Artificial Intelligence", Semantic Scholar , pp. 95-108, June 2019, doi: 10.1007/978-981-13-8285-7_8.

[66] G. Naidu, T. Zuva, and E. M. Sibanda, "A Review of Evaluation Metrics in Machine Learning Algorithms," in Artificial Intelligence Application in Networks and Systems, vol. 724, Lecture Notes in Networks and Systems, Springer, 2023, pp. 15-25. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-35314-7_21.

[67] Hernández-Contreras, D. A., Peregrina-Barreto, H., Rangel-Magdaleno, J. D. J., & Orihuela-Espina, F. (2019). Statistical Approximation of Plantar Temperature Distribution on Diabetic Subjects Based on Beta Mixture Model. IEEE Access, 7, 28383-28391. doi:10.1109/ACCESS.2019.2903346.

[68] G. Naidu, T. Zuva, and E. M. Sibanda, "A Review of Evaluation Metrics in Machine Learning Algorithms," in Artificial Intelligence Application in Networks and Systems, vol. 724, Lecture Notes in Networks and Systems, Springer, 2023, pp. 15-25.

[69] R. K. Gudivaka, R. L. Gudivaka, B. R. Gudivaka, D. K. R. Basani, S. H. Grandhi, and F. Khan, "Diabetic foot ulcer classification assessment employing an improved machine learning algorithm," *Technology and Health Care*, early access, Jan. 19, 2025, doi: 10.1177/09287329241296417.

[70] Sh. Hong, Y. Chen, Y. Lin, X. Xie, G. Chen, and H. Xie, "Personalized prediction of diabetic foot ulcer recurrence in elderly individuals using machine learning paradigms," Technology and Health Care, vol. 32, pp. S265-S276, 2024, doi: 10.3233/THC-248023

[71] M. Alzyoud, R. Alazaidah, M. Aljaidi, G. Samara, M. H. Qasem, M. Khalid, and N. Al-Shanableh, "Diagnosing diabetes mellitus using machine learning techniques," International Journal of Data and Network Science, vol. 8, pp. 179-188, 2024.

[72] B. Antal and A. Hajdu, "An ensemble-based system for automatic screening of diabetic retinopathy," Knowledge-Based Systems, vol. 60, pp. 20-27, 2014. [Online]. Available:https://archive.ics.uci.edu/ml/datasets/Diabetic+Retinopathy+Debrecen+Data+Set.

[73] Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality. [Online]. Available: https://hcup-us.ahrq.gov/tech_assist/citations.jsp.

[74] J. Zhang and P. Wang, "A new method of color image segmentation based on intensity and hue clustering," in Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain, 2000, pp. 613-616.

[75] J. G. Cleary and L. E. Trigg, "K*: An Instance-based Learner Using an Entropic Distance Measure," in Proceedings of the 12th International Conference on Machine Learning, Tahoe City, CA, USA, 1995, pp. 108-114.

[76] G. H. John and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers," in Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, 1995, pp. 338-345.

[77] N. Landwehr, M. Hall, and E. Frank, "Logistic Model Trees," Machine Learning, vol. 59, no. 1-2, pp. 161-205, 2005.

[78] S. Stefanopoulos, Q. Qiu, G. Ren, A. Ahmed, M. Osman, F. C. Brunicardi, and M. Nazzal, "A Machine Learning Model for Prediction of Amputation in Diabetics," Journal of Diabetes Science and Technology, vol. 18, no. 4, pp. 874-881, 2024, doi: 10.1177/19322968221142899.

[79] National Inpatient Sample (NIS) database from 2008 to 2014," Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality, Rockville, MD, USA. Available: https://www.hcup-us.ahrq.gov/nisoverview.jsp.

[80] N. Arteaga-Marrero, A. Hernández-Guedes, J. Ortega-Rodríguez and J. Ruiz-Alzola, " State-of-the-Art Features for Early-Stage Detection of DF Ulcers Based on Thermograms", Biomedicines, vol. 11, no. 12, pp. 1-16, Nov 2023, doi: https://doi.org/10.3390/ biomedicines11123209.

[81] D. Bouallal, A. Bougrine, R. Harba, R. Canals, H. Douzi, L. Vilcahuaman and H. Arbanil, "STANDUP database of plantar foot thermal and RGB images for early ulcer detection," Open Research Europe, vol. 2, no. 77, Jun. 2022.

[82] D. A. Hernandez-Contreras, H. Peregrina-Barreto, J. de Jesus Rangel-Magdaleno and F. J. Renero-Carrillo, "Plantar thermogram database for the study of DF complications," IEEE Access, vol. 7, pp. 161296-161307, 2019, doi: https://ieeexplore.ieee.org/document/8883670.

[83] N. Arteaga-Marrero, L. C. Bodson, A. Hernández, E. Villa and J. Ruiz-Alzola, "Morphological Foot Model for Temperature Pattern Analysis Proposed for DF Disorders," Applied Sciences, vol. 11, no. 16, p. 7396, Aug. 2021, doi: https://www.mdpi.com/2076-3417/11/16/7396.

[84] V. Filipe, P. Teixeira, and A. Teixeira, "Automatic Classification of Foot Thermograms Using Machine Learning Techniques," Algorithms, vol. 15, no. 7, pp. 236, 2022, doi: 10.3390/a15070236.

[85] A. Khandakar, M. Chowdhury, M. Reaz, S. Ali, S. Kiranyaz, T. Rahman, M. H. Chowdhury, M. Ayari, R. Alfkey, A. Bakar, R. Malik, and A. Hasan, “A Novel Machine Learning Approach for Severity Classification of DF Complications Using Thermogram Images”, Sensors, vol. 22, no. 11, 2022, doi: 10.3390/s22114249.

[86] V. Sudha and D. Ganeshbabu, “A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning”, Comput. Mater. Contin., vol. 66, pp. 827–842, 2020, doi: 10.32604/cmc.2020.012008.

[87] A. Natekin and A. Knoll, “Gradient Boosting Machines, A Tutorial”, Front. Neurorobot., vol. 7, p. 21, 2013, doi: 10.3389/fnbot.2013.00021.

[88] S. Malik, R. Harode, and A. Singh, “XGBoost: A Deep Dive into Boosting (Introduction Documentation)”, Feb 2020, doi: 10.13140/RG.2.2.15243.64803.

[89] R. P. Mahapatra, S. Kumar, P. Sudip, R. Pritee Parwekar, and L. Goel, “Analyzing Effectiveness of ML Algorithms Used to Predict Diabetes Mellitus Archit”, in Lecture Notes in Networks and Systems 341 Proceedings of International Conference on Recent Trends in Computing, pp. 667–672, 2021.

[90] V. Chang, J. Bailey, Q. A. Xu, and Z. Sun, “Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms.”, Neural Comput. Appl., pp. 1–17, Mar. 2022, doi: 10.1007/s00521-022-07049-z.

[91] A. R. Naidu and D. Bhavana, “Infrared thermal imaging for automatic detection of diabetes using morphological processing”, Int. J. Syst. Assur. Eng. Manag., International Journal of Systems Assurance Engineering and Management, 2021, doi: 10.1007/s13198-021-01294-6.

[92] MathWorks, Inc., “Image Processing Toolbox: MATLAB,” 2021, Available: [online], last visited on 17-Jan-2024.

[93] J. Guzaitis, A. Kadusauskiene, and R. Raisutis, “Algorithm for automated foot detection in thermal and optical images for temperature asymmetry analysis,” Electron., vol. 10, no. 5, pp. 1–14, 2021, doi: 10.3390/electronics10050571.

[94] https://businessdesk.co.nz/settings#/subscription, last visited on 12/1/2024.

[95] A. Khandakar, M. Chowdhury, M. Ibne Reaz, S. Md Ali, M. Hasan, S. Kiranyaz, T. Rahman, R. ALfkey, A. Bakar, and R. Malik, “A machine learning model for early detection of DF using thermogram images”, Comput. Biol. Med., vol. 137, no. September, p. 104838, 2021, doi: 10.1016/j.compbiomed.2021.104838.

[96] D. Elreedy and A. Atiya, " A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance", Information Sciences, vol. 505, pp. 32-64, Dec 2019, doi: https://doi.org/10.1016/j.ins.2019.07.070.

[97] J. Saminathan, M. Sasikala, V. B. Narayanamurthy, K. Rajesh, and R. Arvind, “Infrared Physics & Technology Computer aided detection of DF ulcer using asymmetry analysis of texture and temperature features”, Infrared Phys. Technol., vol. 105, no. January, p. 103219, 2020, doi: 10.1016/j.infrared.2020.103219.

[98] C. Headquarters, “FLIR E50 (incl. Wi-Fi),” FLIR Systems, Inc., Wilsonville, OR, USA, Tech. Rep., 2012, doi: https://doi.org/10.13140/RG.2.2.19592.89603.

[99] P. L. Li, Q. F. Xiao, K. L. Yick, Q. L. Liu, and L. Y. Zhang, “A novel deep learning approach to classify 3D foot types of diabetic patients,” Sci. Rep., vol. 15, no. 1, p. 13819, Apr. 2025, doi: 10.1038/s41598-025-98471-5.

[100] https://www.shining3d.com/professional-solutions/multifunctional-handheld-3d-scanners/einscan-pro-hd, last visited on 25-5-2525.

[101] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. 31st Int. Conf. Neural Information Processing Systems (NIPS)*, Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 5105–5114

[102] V. Panamonta, R. Jerawatana, P. Ariyaprayoon, P. Looareesuwan, B. Ongphiphadhanakul, C. Sriphrapradang, L. Chailurkit, and B. Ongphiphadhanakul, "Plantar thermogram analysis using deep learning for diabetic foot risk classification," J. Diabetes Sci. Technol., early access, 2025. [Online]. Available: https://doi.org/10.1177/19322968251316563.

[103] L. Z. Chee, S. Sivakumar, K. H. Lim, and A. A. Gopala, “Gait acceleration-based diabetes detection using hybrid deep learning,” Department of Electrical and Computer Engineering, Curtin University Malaysia, CDT 250, 98009, Miri, Sarawak, Malaysia, and Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway, 47500, Subang Jaya, Selangor, Malaysia.

[104] A. M. El-Kady, M. M. Abbassy, H. H. Ali, and F. A. Moussa, “Advancing Diabetic Foot Ulcer Detection Based on ResNet and GAN Integration,” Journal of Theoretical and Applied Information Technology, vol. 102, no. 6, pp. 2258-2268, Mar. 2024.

[105] Gulshan and A. Sh. Arora, " Automated prediction of diabetes mellitus using infrared thermal foot images: recurrent neural network approach", Biomedical Physics & Engineering Express, Vol. 10, No. 2, 12 February 2024, doi: 10.1088/2057-1976/ad2479.

[106] M. H. Alshayeji, S. Ch. Sindhu, and S. Abed, " Early detection of DF ulcers from thermal images using the bag of features technique", Biomedical Signal Processing and control, vol. 79, pp. 104143, Jan 2023, doi: https://doi.org/10.1016/j.bspc.2022.104143.

[107] S. Gamil, M. Rahman and A. Haider, " Bag of Features (BoF) Based Deep Learning Framework for Bleached Corals Detection", Big Data and Cognitive Computing, vol. 5, no. 4, pp. 1-15, 2021, doi: https://www.mdpi.com/2504-2289/5/4/53.

[108]T. Lindeberg, "Scale Invariant Feature Transform", Digital Image Processing: An Algorithmic Introduction, vol. 7, 2012, doi: 10.4249/scholarpedia.10491.

[109]H. Bay, A. Ess, T. Tuytelaars and L. V. Gool, " Speeded-Up Robust Features (SURF)", in Computer Vision – ECCV 2006. LNCS, vol. 110, no. 3, pp. 346-359, June 2008, doi: https://doi.org/10.1016/j.cviu.2007.09.014.

[110]R. N. Yousef, M. M. Eid, M. A. Mohamed, " Classification of DF Thermal Images Using Deep Convolutional Neural Network", Journal of Intelligent Systems & Internet of Things , vol. 8, no.1, pp. 17-32, 2023, doi: 10.54216/JISIoT.080102.

[111]J. Reyes-Luévano, J.A. Guerrero-Viramontes, J. Rubén Romo-Andrade and M. Funes-Gallanzi, " DFU_VIRNet: A novel Visible-InfraRed CNN to improve DF ulcer classification and early detection of ulcer risk zones", Biomedical Signal Processing and Control, vol. 86, pp. 105341, Sep 2023, doi: https://doi.org/10.1016/j.bspc.2023.105341.

[112]V. Divol, J. Niles-Weed, and A.-A. Pooladian, "Optimal transport map estimation in general function spaces," arXiv preprint, pp. 68, DEC 2022, doi: https://doi.org/10.48550/arXiv.2212.03722.

[113]N. Sharma, S. Mirza, A. Rastogi and P.K. Mahapatra, " Utilizing Mask R-CNN for Automated Evaluation of DF Ulcer Healing Trajectories: A Novel Approach", Traitement du Signal, vol. 40, no. 4, pp. 1601-1610, Aug 2023, doi: https://doi.org/10.18280/ts.400428.

[114]H. Shao, "An Improved Vision Transformer for Early Detection of DF Using Thermogram", 2023 8th International Conference on Intelligent Computing and Signal Processing (ICSP), n0. 23746752, April 2023, doi: 10.1109/ICSP58490.2023.10248732.

[115]Sh. Sh. Reddy, L. Alluri, M. Gadiraju and R. Devareddi, " Forecasting DF Ulcers Using Deep Learning Models", Proceedings of Third International Conference on Sustainable Expert Systems: ICSES 2022, pp. 222-238, 2023, doi: https://doi.org/10.1007/978-981-19-7874-6_16.

[116]M. Ahsan, S. Naz, R. Ahmad and A. Sikandar, " A Deep Learning Approach for DF Ulcer Classification and Recognition", Jor. Information, vol. 14, no. 1, Jan 2023, doi: 10.3390/info14010036.

[117]A. Hernandez-Guedes, N. Arteaga-Marrero, E. Villa, G. M. Callico and J. Ruiz-Alzola, "Feature Ranking by Variational Dropout for Classification Using Thermograms from DF Ulcers", Sensors, vol. 23, no. 2, pp. 757, Jan 2023, doi: 10.3390/s23020757.

[118]P. N. Thotad, G. R. Bharamagoudar, and B. S. Anami, "DF ulcer detection using deep learning approaches", Sensors Int., vol. 4, pp. 100210, 2023, doi: https://doi.org/10.1016/j.sintl.2022.100210.

[119]G. Menghani, "Efficient deep learning: A study on making deep learning models smaller, faster, and better", ACM Computing Studys, vol. 55, no. 12, p. 1-37, 2023, doi: https://doi.org/10.1145/3578938.

[120]V. Khullar, R. G. Tiwari, A. K. Agarwal and M. Angurala, " Transferring Pre-trained Deep CNNs on Plantar Thermograms for DF Disease", Mobile Radio Communications and 5G Networks: Proceedings of Third MRCN 2022, Springer, vol. 588, pp. 120-130, Feb 2023, doi: https://doi.org/10.1007/978-981-19-7982-8_9.

[121]I. Khosa, A. Raza, M. Anjum, W. Ahmad and S. Shahab, "Automatic DF Ulcer Recognition Using Multi-Level Thermographic Image Data", Diagnostics, vol. 13, no. 16, pp. 2637, Aug 2023, doi: 10.3390/diagnostics13162637.

[122]M. Ray, T. Maity, and R. Sharma, "Early detection of foot ulceration in type II diabetic patient using registration method in infrared images and descriptive comparison with deep learning methods", J. Supercomput., The Journal of Supercomputing, vol. 78, no. 0123456789, pp. 1–18, July 2022, doi: 10.1007/s11227-022-04380-z.

[123]A. Khan and S. Safdar, " Symmetric and asymmetric analysis of trade and environment in Pakistan", National Library of Medicine, vol. 5, pp. 11399-11416, 30 Jan 2023, doi: 10.1007/s11356-022-22669-7.

[124]Kh. Munadi, K. Saddami, M. Oktiana, R. Roslider, K. Muchtar, M. Melinda, R. Muharar, M. Syukri, T. F. Abdin, and F. Arnia, "A Deep Learning Method for Early Detection of DF Using Decision Fusion and Thermal Images", Appl. Sci., vol. 12, no. 15, 2022, doi: 10.3390/app12157524.

[125]Sh. Muralidhara, A. Lucieri, A. Dengel, and S. Ahmed, "Holistic multi-class classification & grading of DF ulcerations from plantar thermal images using deep learning", Heal. Inf. Sci. Syst., vol. 10, no. 1, pp. 1–9, 2022, doi: 10.1007/s13755-022-00194-8.

[126]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, doi: 10.1109/ACCESS.2022.3180036.

[127]H. Maldonado, R. Bayareh, I. A. Torres, A. Vera, J. Gutiérrez, and L. Leija, "Infrared Physics & Technology Automatic detection of risk zones in DF soles by processing thermographic images taken in an uncontrolled environment", Infrared Phys. Technol., vol. 105, no. January, p. 103187, 2020, doi: 10.1016/j.infrared.2020.103187.

[128]K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.322

[129]https://sonycine.com/articles/sony-electronics-announces-new-venice-extension-system-2-and-venice-2-version-2-00-firmware-update-with-powerful-new-features-and-capabilities/ , last visited on 12/1/2024.

[130]J. Haglund, J. Fredrik, E. Melander, A.M. Pendrill, C. Xie, and K. Schönborn, "Infrared Cameras in Science Education", Infrared Phys. Technol., vol. 75, 2016, doi: 10.1016/j.infrared.2015.12.009.

[131] B. Ploderer, R. Brown, L. S. D. Seng, P. A. Lazzarini, and J. J. van Netten, “Promoting self-care of DF ulcers through a mobile phone app: User-centered design and evaluation,” JMIR Diabetes, vol. 3, no. 4, Oct. 2018, doi: https://doi.org/10.2196/10105.
[132] S. Assistant and T. Nadu, “Inspection and Detection of DF Using Image”, Int. Res. J. Glob. Eng. Sci., vol. 4, no. 4, pp. 9–13, 2020.
[133] M. Omran, A. Engelbrecht, and A. Salman, “An overview of clustering methods”, Intell. Data Anal., vol. 11, pp. 583–605, 2007, doi: 10.3233/IDA-2007-11602.
[134] J. Roerdink and A. Meijster, “The Watershed Transform: Definitions, Algorithms and Parallelization Strategies”, FUNDINF Fundam. Inform, vol. 41, 2000, doi: 10.3233/FI-2000-411207.
[135] P. Sahoo, S. Soltani, and A. Wong, “A Study of Thresholding Techniques”, Comput. Vision, Graph. Image Process., vol. 41, pp. 233–260, 1988, doi: 10.1016/0734-189X(88)90022-9.
[136] B. Ryabko, J. Astola, and M. Malyutov, "Compression-Based Methods of Statistical Analysis and Prediction of Time Series", ResearchGate, pp. 1-44, Jan 2016, doi: 10.1007/978-3-319-32253-7.
[137] D. Bouallal, A. Bougrine, H. Donzi, R. Harba, R. Canals, L. Vilcahuaman, and H. Arbanil, “Segmentation of plantar foot thermal images: Application to DF diagnosis”, Int. Conf. Syst. Signals, Image Process., vol. 2020-July, pp. 116–121, 2020, doi: 10.1109/IWSSIP48289.2020.9145167.
[138] N. Siddique, S. Paheding, C. P. Elkin, and V. Devabhaktuni, " U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications", IEEE Access, vol. 9, pp. 82031-82057, 2020, doi: 10.1109/ACCESS.2021.3086020.
[139] https://collections.nlm.nih.gov/catalog/nlm:nlmuid-101157238-bk, last visited on 15/4/2023.
[140] L. Alzubaidi, M. A. Fadhel, S. R. Oleiwi, O. Al-Shamma, and J. Zhang, “DFU_QUTNet: DF ulcer classification using novel deep convolutional neural network”, Multimed. Tools Appl., vol. 79, no. 21–22, pp. 15655–15677, 2020, doi: 10.1007/s11042-019-07820-w.
[141] C. R. Knight and C. Winship, “The Causal Implications of Mechanistic Thinking: Identification Using Directed Acyclic Graphs (DAGs),” in Handbook of Causal Analysis for Social Research, S. L. Morgan, Ed. Dordrecht: Springer, 2013, pp. 273-302, doi: 10.1007/978-94-007-6094-3_14.DAG.
[142] N. Alsharman and I. Jawarneh, “GoogleNet CNN neural network towards chest CT-coronavirus medical image classification”, J. Comput. Sci., vol. 16, no. 5, pp. 620–625, 2020, doi: 10.3844/JCSSP.2020.620.625.
[143] I. N. G. Astawa, M. Radhitya, I. W. Ardana, and F. Dwiyanto, “Face Images Classification using VGG-CNN”, Knowl. Eng. Data Sci., vol. 4, p. 49, 2021, doi: 10.17977/um018v4i12021p49-54.
[144] G. G. A. Celano, “A ResNet-50-based Convolutional Neural Network Model for Language ID Identification from Speech Recordings”, SIGTYP 2021 - 3rd Work. Res. Comput. Typology Multiling. NLP, Proc. Work., pp. 136–144, 2021, doi: 10.18653/v1/2021.sigtyp-1.13.
[145] M. Goyal and S. Hassanpour, “A Refined Deep Learning Architecture for DF Ulcers Detection”, Computer Vision and Pattern Recognition, pp. 1–8, 2020, doi: https://doi.org/10.48550/arXiv.2007.07922.
[146] M. Kayalvizhi and D. Maheswari, “a Hybrid Deep Learning Algorithms for Diabetes Mellitus Prediction Using Thermal Foot Images”, Eur. J. Mol. Clin. Med., vol. 7, no. 11, pp. 5176–5183, 2020.
[147] I. Cruz-Vega, D. Hernandez-Contreras, H. Peregrina-Barreto, J. Rangel-Magdaleno, and J. M. Ramirez-Cortes, “Deep Learning Classification for DF”, National Library of Medicine, no. Dm, pp. 1–22, 2020, doi: 10.3390/s20061762.
[148] M. Goyal, N. D. Reeves, A. K. Davison, S. Rajbhandari, J. Spragg, and M. H. Yap, “DFUNet: Convolutional Neural Networks for DF Ulcer Classification,” IEEE Trans. Emerg. Top. Comput. Intell., vol. 4, no. 5, pp. 728–739, 2018, doi: 10.1109/tetci.2018.2866254.
[149] https://trc.nist.gov/ , last visited on 10/4/2024.

Appendix

ATD Absolute Temperature Difference
AUC Area Under the Curve
BN Batch Normalization
CAD Computer Aided Detection
DF Diabetic Foot
DME Diabetic Macular Edoema
GNA Gaussian Naive Bayes
IRT Infrared Thermography
LCA Leteral Callcaneal Artery
LPA Leteral planter Artery
MCA Medial Callcaneal Artery
MPA Medial planter Artery
MRI Magnetic Resonance Imaging
PCA Principal Component Analysis
QDA Quadratic Discriminant Analysis
RELU Rectified Linear Unit
ROI Region Of Interest
TCI Thermal Change Index