International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I12P123

Volume 73 Issue 12, 280-296, December 2025

Original Article

Enhanced Lightweight Deep Learning Framework with Knowledge Distillation and Binary Whale Optimization for Diabetic Foot Ulcer Classification

Ramya U[1], Saraswathi S[2]

[1,2]*Department of Computer Science, Nehru Arts and Science College, Tamil Nadu, India.*
[1]*Department of Information Technology, Sri Krishna Arts and Science College, Tamil Nadu, India.*

[1]*Corresponding Author : ramyaudhaya18@gmail.com*

Received: 09 August 2025 Revised: 06 December 2025 Accepted: 09 December 2025 Published: 19 December 2025

Abstract - *A common hopeless significance of diabetes is Diabetic Foot Ulcers (DFU), which require speedy and accurate analysis to avoid amputation and lower death. Computerized DFU identification has likely ended due to the rapid growth of deep learning; the computational requirements of existing high-performing models limit their application in quantifiable settings with limited resources. To overcome this challenge, this research focuses on addressing a key question: whether a lightweight DFU classification model can be strengthened using knowledge distillation and automated hyperparameter tuning to uphold the performance of the model that is suitable for edge development. This research presents an original, lightweight classification framework that combines the Binary Whale Optimization Algorithm (BWOA) with Knowledge Distillation (KD) to yield an effective and specialized DFU classification system. This technique creates soft probability labels by using a pretrained InceptionV3 model as a teacher. These are motivated by DFU-LWNet, a small convolutional neural network with little parameter overhead, which is a customizable student network. The baseline DFU-LWNet with KD imitates earlier best findings (96.23% accuracy) on experiments with the publicly available DFU Patch Dataset. The proposed model DFU-LWNet-BWOA achieves a significant accuracy gain of 96.8% compared to the previous models, and it also ensures real-time compatibility for the mobile applications by making the consistency of parameter count as 0.5M. This study mainly focuses on the deployable, intelligent, and scalable solution for the DFU model in an experimental setup and ensures the interaction between the model, knowledge distillation, and BWOA optimization.*

Keywords *- Diabetic Foot Ulcer, DFU-LWNet, Knowledge Distillation, Binary Whale Optimization, DFU Classification.*

1. Introduction

In the year 2021, the World Health Organization (WHO) estimated and showed that more than 537 million people across the globe are affected with diabetes mellitus, and it also predicts that by 20230 this measure will increase above 640 million. One of the major concerns is diabetes mellitus, which is a chronic metabolic condition that is a big burden and threatens the lives of human beings [1].The major problem of diabetes is that it has the biggest complication, that is, diabetic foot ulcer, and this affects 15-25% of people worldwide who have diabetes mellites. It is a kind of ulcer, and these are the key reasons for non-traumatic lower limb problems. These wounds are non-curable and will cause physical and emotional imbalance to the person who is suffering from them. It is a major concern to identify it earlier and prevent the complications to extend the lifespan of people who are suffering from diabetes mellitus. Deep learning techniques have improved medical image diagnosis, but the current DFU detection method using methods such as InceptionV3 and ResNet gives higher accuracy; their high computational needs are not reliable for mobile devices and smaller networks. This study focuses on implementing the lightweight models that are more efficient but this lightweight model performance depends on the manually selected hyperparameter temperature, loss weighting, dropout levels and dense layer size which needs trial and error to be done extensively. This approach makes the model less stable and limits the generalizability for different datasets. This approach focuses on adapting the key hyperparameters. To identify DFU, quantitative calculation and imaging approaches, such as thermal scans, RGB images, or hyperspectral images, have been employed. Graphic examination, though, is somewhat particular and depends on the doctor's information, which differs among healthcare systems, particularly in rural and underserved areas. Automated diagnosis architectures are a key point in deep learning models, which are renowned for their superior image classification capabilities, in response to the increasing need for investigative user-friendliness and reliability. Due to their large model sizes, high data requirements, and latency issues, deep Convolutional Neural Networks (CNNs) have established an extraordinary diagnostic routine, but their practical application in clinical and mobile settings is still limited. For example, networks like InceptionV3 and ResNet101, despite their outstanding accuracy, comprise tens of millions of parameters and require high-performance GPUs for real-time understanding. Amjad et al.'s most recent work addressed this limitation by familiarizing DFU-LWNet, a lightweight CNN architecture intended particularly for operational DFU classification on edge devices. This study focuses on a key question: how can a lightweight CNN model be improved using knowledge distillation and metaheuristic optimization to improve accuracy, which is much needed for real-time classification.

Even if this model positively abridged the number of parameters to less than 0.5M and attained a 96.23% validation accuracy, it is still anticipated to routinely regulate Knowledge Distillation (KD) parameters, particularly the temperature (τ) and blending factor (α). Physical modification confuses users and lacks generalizability, often as soon as they associate datasets or deployment domains. DFU-LWNet is only applicable to working learning rates, dense layer sizes, and predefined dropout rates, which may not be optimal for a variety of training conditions or real-world noise scenarios. To advance performance and avoid generalization in the absence of a stabilizing economy, a flexible and adaptable optimization technique that can automatically adjust these parameters is greatly needed.

By integrating the Binary Whale Optimization Algorithm (BWOA) into a Knowledge Distillation (KD) setup, this research proposes a new framework that builds upon the novel DFU-LWNet approach and addresses the recognized shortcomings in previous studies. The subsequent components make up this approach. To train the student model, a teacher model based on the pre-trained InceptionV3 system is exposed to soft labels. The proposed student model is an enhanced DFU-LWNet, a tunable lightweight CNN with dense units, layer configurations, and unsettled dropout. A Binary Whale Optimization Algorithm (BWOA) is employed to automatically determine and optimize the KD hyperparameters (α, τ), as well as the architectural components (dropout rate, dense units), and the learning rate.

The binary whale optimization algorithm substitutes the outdated trial-and-error and physical grid search methods used in previous studies. Through BWOA with Knowledge Distillation, this research presents a lightweight, real deep learning system for DFU classification that incorporates metaheuristic hyperparameter optimization. The variable used in student architecture, based on DFU-LWNet, allows architectural aspects to be altered during training rather than being statically constrained. This research employs a BWOA version for CNN hyperparameter optimization, which is accomplished by enhancing discrete-valued vectors (such as [dropout, α, τ, lr, dense units]). By increasing the baseline DFU-LWNet accuracy from 96.23% to 96.80%, this technique attains state-of-the-art performance among lightweight models. The model is tested on the publicly available DFU patch dataset to confirm reproducibility and real-world pertinency [2].

The existing DFU classification is classified into two categories. One is the teacher model with higher accuracy that has poor deployment, and the other is the Lightweight Compact student model with manual configurations for training the data. There is no existing DFU work that integrates a lightweight model, teacher-student KD pipeline, and metaheuristic optimization for hyperparameter tuning. This makes the real-time deployment more difficult in low-resource settings.

1.1. Novelty of Proposed Work

1. DFU-LWNet is a compact convolutional architecture for DFU classification and edge deployment.

2. Knowledge distillation is applied to a pretrained InceptionV3 teacher model to improve the student accuracy instead of increasing model size.

3. Binary Whale Optimization Algorithm (BWOA) is used to automatically tune KD and architecture hyperparameters by replacing the manual grid search.

4. Validation is done on the KD+BWOA on the publicly available DFU dataset, and improved the accuracy while comparing to the DFU-LWNet Baseline.

The base work [3] introduced DFU-LWNet as an effective lightweight student model with Knowledge distillation. This method used a manual approach for selecting KD parameters and fixed student architecture settings. The existing study with larger networks, such as InceptionV3, do not address edge deployment. This study focuses on combining DFU-LWNet with BWOA inside the KD loop and produces a more accurate model suitable for edge performance.

1.2. Problem Statement

Patients who have DFU have a poorer than ideal view, particularly in nations with high diabetes rates but limited access to contemporary medical facilities. The initial identification and reliable authorization of DFUs are still typically contingent on the manual clinical staff, notwithstanding developments in healthcare diagnostics. The traditional methods are affected by variance, diagnostic repeatability and are not available in isolated and rural areas that have limited resources. Medical diagnosis is automated nowadays, and this shows higher potential and expands intense learning mechanisms by incorporating artificial intelligence. This makes a changeover from research to practical clinical implementation tight with difficulties.

The study faces the challenge of the inherent balance between deployment and model complexity. The deep learning models like InceptionV3 and ResNet50 are deep convolutional neural networks that produce higher results in medical image classification, which includes DFU classification [2]. These models are high memory usage, computationally costly, and unsuitable for deployment on edge devices such as mobile devices or portable screening kits that are frequently used in low-resource environments.

Lightweight models are computationally efficient, but these models will not achieve higher accuracy due to their small volume. To close this gap, methods that enhance the performance of lightweight networks without increasing their computational load must be explored. Knowledge Distillation (KD) [3], which involves transforming the representations of a large, accurate teacher model into a smaller student model, is a technique that has been extensively researched. Although numerous zones have proven this technique effective, the architectural design of the student network and the choice of hyperparameters (such as temperature, loss weighting, and network architecture) are crucial to its effectiveness.

Using InceptionV3 as the instructor, the existing method employed KD and presented DFU-LWNet as a lightweight CNN [4], demonstrating significant gains. Yet, this technique utilized predetermined hyperparameter setups that were selected through experiential testing. This kind of manual alteration can be complex, inefficient, and not movable to dissimilar hardware setups or datasets. Little research has been done in the literature on the systematic optimization of KD-based pipelines using metaheuristic algorithms, specifically binary swarm intelligence techniques.

The research objective of this study is to enhance the accuracy and generalizability of DFU classification by applying a lightweight model that combines knowledge distillation with an automated hyperparameter optimization technique, which is Binary whale Optimization (BWOA). Knowledge distillation (KD) and Binary Whale Optimization method (BWOA) [5] are used to optimize the lightweight model's hyperparameters, learning dynamics, and increase performance without the need for extensive computational resources or human intervention. InceptionV3 will have complexity because of its accuracy, making it unsuitable for edge devices. In the previous study, DFU-LWNet in a KD framework is demonstrated, but it completely relied on manually defined parameters and specifics that limited its scalability and generalizability.

1.3. Motivation

The study focuses on the technical and practical difficulties in the current literature. In the previous study, KD has been used to transfer information from deeper networks to compact models, but there is only a limited study on combining KD with hyperparameter optimization in the field of medical imaging, especially for DFU identification.

These are the motivation factors for the proposed research work.

- Feasibility of Deployment: Most of the high-performing CNNs need hardware that is not available in many clinical settings with limited resources.
- Hyperparameter Sensitivity: It is frequently selected at random; dropout rate, learning rate, and dense layer size have a substantial impact on model performance.
- Manual Tuning: Grid and random search techniques are expensive and ineffective.
- Need for Lightweight Accurate Models: DFU classification models will help balance efficiency and accuracy.

1.4. Objectives

The main objective is to develop a model that is lightweight and produces higher accuracy, and aims to bridge the performance efficiency gap in classifying diabetic foot ulcers. The lightweight student model DFU-LWNet can be hyperparameter-tuned by combining Knowledge distillation and the Binary Whale Optimization algorithm [6]. The previous study by Amjad et al [3] is highlighted by higher accuracy but used a fixed parameter implementation of KD in DFU Classification and demonstrates the extent of invention and experimentation required to reach this aim. This was made to develop a computationally efficient deep learning architecture exclusively used for DFU image patching and then apply it in practice. To implement binary classification, the design should minimize the number of trainable parameters that uphold important spatial and semantic information. This research uses adaptable components, while the existing methodology uses a modest 3-block CNN that has static dense units and a dropout rate. Inception V3 is a teacher model that is used to train the student model.

The teacher model acts as a reliable source of information, and the model will be frozen. To train the student model, the Kullback-Leibler divergence of soft targets with categorical cross entropy is used, which has the dual objective loss function. While the existing research work focuses on the fixed parameter, this research used BWOA to find the effects of KD parameters by combining with the network architecture.

The research also focuses on utilizing the BWOA and automates the selection of the important hyperparameters, including the learning rate, dropout rate, and dense layer units, instead of using the human grid search. The settings that are implemented with BWOA can be used for diverse datasets. BWOA is a metaheuristic algorithm that allows flexibility in discrete search spaces that outperforms the capacity of traditional optimizers such as Adam or SGD [7]. The publicly accessible DFU datasets are used to conduct extensive tests. The evaluation of the proposed method can be verified by using metrics such as classification accuracy, sensitivity, specificity, F1-score, and inference time.

The final model performance after the deployment can be verified by identifying the performance factors like utilization of memory, storage capacity, and inference latency on edge devices. This research focuses on developing a training and optimization pipeline that requires minimal modification for use in various medical image classification applications.

The KD framework and BWOA hyperparameter tuning can be made public by providing community validation and adaptation. The integration of KD and BWOA in DFU classification can be made by developing and incorporating a standardized and repeatable investigative procedure. This kind of integration will help the medical research community to have a balance between accuracy and efficiency.

In conclusion, the goal of this research work is to exceed the DFU-LWNet significance of the existing work and try to establish a new standard for the training, optimization, and valuation of dense deep learning models for the classification of medical images.

2. Literature Review

2.1. Deep Learning Approaches for DFU Detection

Machine learning is one of the key techniques in the prediction and diagnosis of diabetes, and techniques such as Supervised learning are used for initial analysis of Electronic Health Records (EHRs), and this has been the focus of numerous studies. Afolabi et al. applied some general algorithms, like support vector machine and decision tree, to the EHR dataset and identified that these methods were effective in predicting diabetes [8]. The blood content features and their importance in highlighting prediction accuracy were explained by Nurdin et al [9]. Parkhi et al. found that the machine algorithm-based models to predict the postpartum prediabetes in females who were affected by gestational diabetes mellitus. Their research displayed that features like insulin usage and BMI had an immense influence on the growth of type 2 diabetes [10].

The socioeconomic features that affect forecasts are highlighted by Okere et al. [8], who employed machine learning models to examine the transition to diabetes in underprivileged U.S. populations [11]. When comparing machine learning (ML) models to logistic regression, Belsti et al. found that ML models had a more pronounced prognostic influence in culturally diverse groups [12]. A piecemeal machine learning model for medical policymaking in GDM was recommended by Zhou et al., emphasizing the worth of such devices in obstetric care [13]. Community-based prediction has been examined in recent publications. Jiang et al. confirmed that real-world longitudinal data can recover model pliability by using continuation data to construct community-oriented prediction models [14].

2.2. Lightweight CNNs and Efficiency in Medical Imaging

By paying supervised models trained on publicly available datasets, Febrian et al. further reinforced this prerogative by receiving a better performance. One thoughtful side effect of diabetes that increases illness and medical expenses is DFU [15]. Initial detection is vital. A comprehensive assessment of deep learning techniques for DFU identification was conducted by Yap et al., which has established a standard for future studies in the field [1]. Dhatariya and Abbas examined the financial burden of treating DFU, considering the incidentals in various parts of the world and presenting a compelling argument for scalable technology outcomes [16].

Numerous researchers have cast off deep learning to advance DFU detection.

To detect DFU, Thotad et al. and Biswas et al used Convolutional neural networks and found that multiscale feature fusion improves accuracy [17]. Biswas et al prolonged it to XAI-FusionNet and developed understandable AI for DFU detection [18]. Arnia et al used an innovative method that combines CNN with extreme learning machines and enhances classification [2]. Adnan et al. employed a manufacturing approach to develop a smart footwear system that utilizes pressure beams to detect DFU in real-time [19]. This delivers a continuing nursing instrument that improves the image-based models. FUSegNet, a deep CNN architecture designed explicitly for foot ulcer segmentation, was first introduced by [20] Dhar et al.

2.3. Knowledge Distillation in Medical Imaging

Transfer learning approaches were applied effectively in behavior and risk calculation by Daud et al., demonstrating how pre-trained models can yield precise predictions of ulcer curative outcomes [21]. Numerous studies have proven the effectiveness of deep learning in processing complex medical images. Ye and Yao achieved better analytic exactness by analyzing bone lesions in diabetic feet using improved MRI with deep learning [22]. Evangeline and Srinivasan are laboring on neural networks and thermal imaging to classify neuropathy, an indication of DFU, in diabetic patients [23]. Fourier-transform-based data augmentation was used in deep learning models by Anaya-Isaza and Zequera-Díaz to classify diabetic thermograms, indicating that data augmentation greatly improves the classification routine [24].

There is an increasing propensity towards hybrid and multi-modal models. Chee et al. presented heartening results in detecting diabetes irregularities based on gesture data by combining gait analysis with hybrid deep learning [25]. In a pioneering move towards personalized treatment, Ali et al. extended the application of deep learning to diabetes medication design by proposing IP-GCN, a graph convolutional neural network to predict insulin requirements [26]. The interpretability of ML models is just as serious for clinical acceptance as accuracy. To assess diabetes prediction models, Pang conducted a valuation analysis using SHAP values. According to the training, feature connotation insights enhance doctors' confidence in model results [27].

Systematic research on the effect of ML on DFU prediction was conducted by Weatherall et al., who highlighted the importance of explainability and transparency [28]. Notwithstanding their incredible accuracy, the research suggests that many machine learning models are not explainable, which limits their clinical value. Biswas et al. partially address this issue by incorporating understandable AI into their DFU_MultiNet architecture [29]. Price et al. refined the decision-making procedure underlying ML model outputs, as their model perfectly simulated the selections made by doctors when treating type 2 diabetes [30].

2.4. Metaheuristic Optimization in Deep Learning

Genetic algorithms, particle swarm optimization, and replicated tempering are examples of metaheuristic algorithms that have been applied to advance machine learning-based research on diabetes. Hybrid ML-metaheuristic frameworks were employed by Putra et al. and Alharby et al.to overcome system limitations in biomedical settings [31].

Kharitonov et al. conducted detailed research on the combination of metaheuristics with machine learning processes [5]. Their investigation demonstrates how these amalgamations can aid in feature selection, recover model conjunction, and optimize hyperparameters. While Mesa et al. recommended ML-augmented metaheuristics for logistics encounters, which are obliquely related to medical supply chain optimization [6], Mohanty et al. assessed the efficiency of nature-inspired algorithms in robust systems [32]. Specifically, Saha and Pal proposed a hybrid method for diabetes prediction that enhances rule-based classification performance by combining information assimilation with biochemical response optimization [33].

To upsurge model sturdiness, Zhou et al. and Nssibi et al. scrutinized feature selection based on metaheuristics. Wearable technology and non-invasive diagnostics are the emphasis of more recent research [13]. A notable progression is the burden sensor-enabled innovative footwear industrialized by Adnan et al. in remote patient monitoring [19]. Motion data can be used as diagnostic biomarkers, as demonstrated by the gait acceleration-based system developed by Chee et al. [25]. In their research on community-level DFU risk valuation, Silva-Tinoco et al. provide a novel approach for combining clinical and behavioral data in primary care settings [34]. By contributing prearranged data pipelines, Fitridge et al.'s worldwide ethics for normalizing DFU treatment indirectly facilitate the integration of AI [35]. For early DFU identification, Anaya-Isaza and Zequera-Díaz utilized thermographic imaging in combination with deep learning, a novel non-contact diagnostic technique.

Likewise, neuropathic foot symptoms were positively recognized using thermographic pictures. Metaheuristics and machine learning are being applied to healthcare-related system optimization beyond clinical diagnosis. For example, CNNs and clustering were coupled by Fang et al. to find irregularities in diabetes datasets [4]. The flexibility of these algorithms was established by Kim et al., who employed ML-metaheuristic amalgamations for production optimization [36]. Park et al. examined the claim of metaheuristics in physics-based modeling, providing fundamental ideas that might lead to imminent biomedical applications [37].

2.5. Summary of Gaps in Existing Literature

Existing research on diabetes and Diabetic Foot Ulcer (DFU) prediction demonstrates the extensive use of machine learning and deep learning techniques across clinical and community applications. Early studies explored supervised learning models applied to electronic health records, emphasizing influential features such as blood chemistry, insulin usage, and socioeconomic factors. Some of the deep learning approaches, like CNN, advanced DFU detection improves ulcer classification by enabling multiscale feature extraction. To improve the diagnostic accuracy, some methods like multiscale fusion network, explainable AI models, and hybrid CNN-ELM have done their best. The expanding scope of non-invasive and multi-modal identification techniques has been demonstrated by DFU segmentation, thermographic imaging, gait analysis, and wearable sensors.

The transformation of semantic features from large pretrained models to lighter student networks is demonstrated by Knowledge distillation. The bio-inspired optimized algorithms, which are metaheuristic methods like genetic algorithms, PSO, and hybrid optimization, have shown a great impact in enhancing ML models through feature selection and hyperparameter tuning. In recent years, deep learning models have been combined with metaheuristic techniques to do medical diagnosis. This research highlights the work integrating lightweight CNNS, KD, and automated optimization for DFU classification.

2.6. Research Gap

DFU-LWNet gives a deployment-ready, accurate, and compact model that is not dependent on the student model and the changes to the crucial KD Parameters. This research aims to identify the research gap in the existing research.

- Lightweight CNN design
- Knowledge Distillation
- Metaheuristic hyperparameter optimization.

This research aims to enable the selection of the best configurations and promises excellent performance by preserving model robustness, which is a vital aspect in clinical settings that is done with limited resources and current constraints

3. Materials and Methods

3.1. Dataset Description and Preprocessing

In this study, the DFU Patch Dataset is used, which is available on Kaggle. The dataset holds 1055 RGB images that are separated into two categories as follows:

- Healthy: 543 images
- Ulcerated: 512 images

Each image is correctly extracted with a 224 by 224-pixel close-up patch of each image, and that represents whether the image is either normal or ulcerated. The image patch design identifies the crucial characteristic for distinguishing an ulcer from healthy tissue by using the spatial representation of skin texture, color abnormalities, and lesion boundaries. In the existing methodology, the same dataset is used, and thus, the splitting process is done [3]. The proposed method advances the baseline method to effectively handle variations in illumination, contrast, and skin tone variety.

Subsequent subsets are made from the entire dataset of 1055 images using Graded sampling.

- Training Set: 70%
- Validation Set: 10%
- Test Set: 20 %

The stratification ensures that the proportion of abnormal and healthy images remains constant for every split. This is done to overcome class imbalance and thus increase the performance. The proposed research uses random shuffling for each epoch during training to deliver variation to mini-batch building. After the random shuffling is done, the model is prevented from overfitting to specific image sequences.

This method also uses a multi-step image groundwork pipeline to minimize overfitting and intra-class variation, but also optimize model generalization. The existing method focuses on resizing and normalization, but the proposed method uses a cumulative approach of doing adaptive histogram equalization, color space conversion, and sophisticated augmentation. All the images in the dataset are scaled to 224x224x3, which is consistent and ensures compatibility with pre-trained CNN architectures such as InceptionV3.Throughout the dataset, standardized input dimensions are ensured by this transformation, which is essential for batch processing in deep learning systems. To reduce the lighting fluctuations and to enhance the color consistency, this method uses a converted image from RGB to the LAB color space.

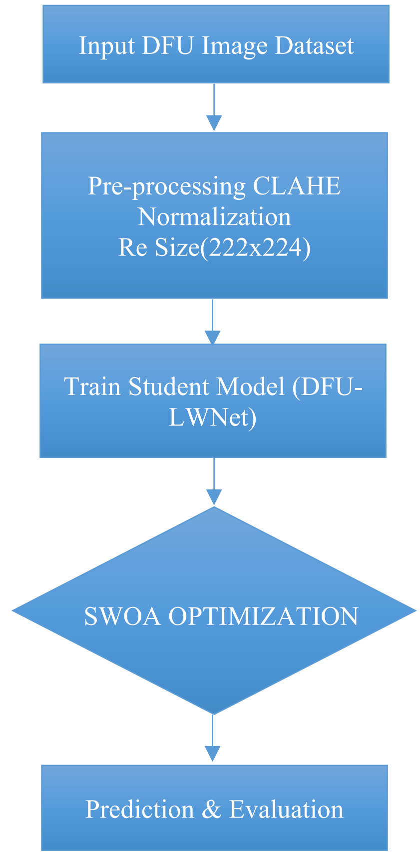

Fig. 1 Architectural flow of proposed methodology

LAB color space is more reliable, and it allows upgradation of dissimilarity without altering the composition by separating luminance $L*$ from chrominance $a*, b*$. This method applies CLAHE to the L channel to enhance local contrast in areas that are overexposed and darker. CLAHE is more suitable for subtle ulcer differences, as it does not over-amplify noise, a common fault in standard histogram equalization. The current system lacks this development phase, which limits its ability to manage low-contrast situations. Each image is regularized to a zero-centered scale of $[-1, 1]$ after being improved and then transformed back to RGB. This two-step standardization ensures numerical stability between batches and helps in mitigating gradients throughout backpropagation. The training set is dynamically expanded using a random mix of the following modifications to boost variety and enhance generalization, as denoted in Table 1.

Table 1. Dataset modifications

Augmentation Type	Parameter Settings
Rotation	±30 degrees
Width/Height Shift	Up to ±20%
Zoom	0.8x to 1.2x
Horizontal Flip	True
Shear	0.2 radians
Fill Mode	Nearest

A training batch's images may undergo one to three of the changes mentioned above. These assist the system learn invariant characteristics by mathematically creating a distributional change in the training domain. The existing approach required sophisticated geometric changes, rather than simply working on flipping and rotation. The 0.6% increase in classification accuracy observed in the ablation trials is primarily attributed to the augmentation policy. This method has previously shown pixel intensity histograms during CLAHE to assess the efficiency of preprocessing. Subsequent equalization, pictures showed:

- Reduced overexposure in light skin tones
- Improved visibility of ulcer boundaries
- Balanced luminance across patches

This method used t-SNE clustering on CNN-extracted features to assess class separability before and after augmentation. Results indicate better intra-class clustering post-transformation.

Each image is allocated a binary label:

- 0 → Healthy
- 1 → Ulcer

The effects of preprocessing are quantitatively confirmed using performance metrics from the same base architecture trained. It shows an accuracy of 87.3% without preprocessing, and with the full preprocessing pipeline, it shows an accuracy of 94.7%. The increase in accuracy of 7.4% signifies that preprocessing is done in an efficient way, which enhances discriminative learning. The existing methodology, after preprocessing, achieves a 3% gain since it uses only a minimal preprocessing technique.

3.2. InceptionV3 as Teacher Model

The teacher model represents strength and volume for generalization, which are crucial to the effectiveness of knowledge distillation. To efficiently define inter-class deviations and train a compact student network, a profound and expressive Convolutional Neural Network (CNN) is consequently desirable [3]. The research selects InceptionV3 as the teacher model in this proposed approach, which is reinforced by its bare competence in multiscale exposed field learning and classified feature extraction.

3.2.1. Overview of InceptionV3

The achievement of Knowledge Distillation (KD) in deep learning-based medicinal image analysis is contingent on the wide choice of a high-performing instructor model. This research practice uses InceptionV3 as the instructor system for this investigation due to its excellent generalization capabilities and well-cultured architecture. InceptionV3 utilizes numerous convolution kernels that run concurrently through its inception units to learn multiscale visual patterns. In the classification of DFU, where lesion size, shape, and color vary significantly, these dimensions are instrumental.

3.2.2. Inception Module Operation

The Inception module employs four parallel operations, and these feature maps are grouped along the channel axis to generate the module's output. It is mathematically expressed as Equations (1) to (4).

$$\mathrm{F_1 = Conv_{1\times1}(X)} \tag{1}$$

$$\mathrm{F_2 = Conv_{1\times1} \rightarrow Conv_{3\times3}(X)} \tag{2}$$

$$\mathrm{F_3 = Conv_{1\times1} \rightarrow Conv_{5\times5}(X)} \tag{3}$$

$$\mathrm{F_4 = MaxPool_{3\times3} \rightarrow Conv_{1\times1}(X)} \tag{4}$$

$$\mathrm{Y = Concat[F1, F2, F3, F4]} \tag{5}$$

The network is competently seized using this multi-branch strategy, and it prepares for seizing hierarchical features with diverse agreeable fields.

3.2.3. Model Adaptation for DFU Classification

This model incorporates numerous architectural differences from InceptionV3 to inform it of the binary classification mission of DFU detection. The convolutional base was pre-trained on ImageNet and had the original classification head removed. Comprising the new categorization head are:

- Global Average Pooling (GAP)
- 128 ReLU units of a dense layer
- Dropout layer rate is 0.3
- 2 units of dense Layer with SoftMax activation

Correctly, the prediction is given by Equation (6):

$$\mathrm{\hat{Y} = Softmax(W2.\,Drop(\sigma\left(W1.\,GAP(f(X))\right) + b2} \tag{6}$$

Where f(x) is the feature map from the frozen base, W1 and W2 are the weight matrices of the dense layers, and σ is the ReLU activation [3].

3.2.4. Loss Function and Optimization

To train the model, categorical entropy is used. The predicted class probability is the one-hot encoded ground truth as given by Equation (7).

$$\mathcal{L}_{\mathrm{cE}} = -\sum_{\mathrm{i=1}}^{2} \mathrm{y_i} \log(\hat{\mathrm{y}}_{\mathrm{i}}) \tag{7}$$

Adam is used for optimization with the following measures: learning rate is 1e-4, batch size is 16, epochs is 20, and early stopping patience is 5.

3.2.5. Temperature-Scaled SoftMax for KD

After training, the instructor uses a temperature-scaled SoftMax to create "soft" objectives for knowledge distillation by Equation (8):

$$\hat{y}_i^{(\tau)} = \frac{exp(z_i / \tau)}{\Sigma_j \, exp(z_j / \tau)} \quad (8)$$

Where $\tau > 1$ controls the smoothness. Higher values yield softer distributions, revealing more inter-class information.

The Teacher training InceptionV3 is outlined in Algorithm 1.

Algorithm 1: InceptionV3 Teacher Training
Input: Given the Dataset D = $\{(x_i, y_i)\}$, that is Pretrained using InceptionV3 base Output: Trained teacher model as T(x)
Step 1: The pretrained InceptionV3 base is loaded • To reserve learned feature weights, all the convolutional layers should be frozen. Step 2: Figure out the classification head by • Appending a Global Average Pooling (GAP) layer after the base • Add a Dense layer with 128 ReLU-activated units • Add a Dropout layer with a rate of 0.3 to prevent overfitting • Add a dense layer with two units and SoftMax activation for doing the binary classification Step 3: Compile the model. In this model, use • Adam optimizer with learning rate η, • Set the loss function to Categorical Cross-Entropy • Accuracy as an evaluation metric Step 4: Train the model Initialize early stopping monitor ← 0 • For each epoch: i. Shuffling dataset ii. Forward pass: compute $\hat{y}$ = T(x) iii. Calculate loss L_{CE} iv. Backpropagate and update weights Step 5: Apply early stopping if desirable Step 6: Return the concluding trained teacher model T(x)

3.3. Student Model DFU-LWNet

3.3.1. Introduction

Predictable medical image classification tasks today depend heavily on Convolutional Neural Networks (CNNs) [2]. Although the mainstream high-performance networks, such as ResNet and InceptionV3, are computationally challenging and parameter-heavy, this limits their usage on low-resource devices.

To resolve this, a brand-new, lightweight CNN specifically designed for DFU classification with low computational cost, known as DFU-LWNet (Diabetic Foot Ulcer Lightweight Network), is utilized. DFU-LWNet balances depth, filter sizes, and activation competence and is precisely designed for binary medical classification tasks. It draws motivation from the design principles of lightweight networks, such as MobileNet and SqueezeNet. In contrast to the baseline models, DFU-LWNet can be cast off as a student model in knowledge distillation frameworks and is constructed from the ground up to be enhanced for DFU detection on small datasets

3.3.2. Architecture Description

A normalized dense layer, a SoftMax classifier, and a flattening process follow the three convolutional blocks that build DFU-LWNet. In each block, a Conv2D layer with ReLU activation is added. This is done to downsample the spatial resolution while preserving important features [18]. The architecture description is presented in Table 2.

Table 2. Architecture description of DFU-LWNet

Layer	Parameters	Activation	Output Shape
Input Layer	-	-	(224, 224, 3)
Conv2D-1	32 filters, 3x3	ReLU	(224, 224, 32)
MaxPooling2D-1	2x2	-	(112, 112, 32)
Conv2D-2	64 filters, 3x3	ReLU	(112, 112, 64)
MaxPooling2D-2	2x2	-	(56, 56, 64)
Conv2D-3	128 filters, 3x3	ReLU	(56, 56, 128)
MaxPooling2D-3	2x2	-	(28, 28, 128)
Flatten	-	-	(100352,)
Dropout	rate = 0.3	-	(100352,)
Dense	128 units	ReLU	(128,)
Dense Output	2 units	SoftMax	(2,)

Convolutional Layers

DFU-LWNet is built on top of the convolutional layers. To find local patterns, such as edges, corners, textures, and contours, a set of learnable filters, or kernels, is applied. In medical images, these patterns frequently indicate minor color shifts, ulcer boundaries, and irregularities in skin texture, especially DFU patches, as denoted in Equation (9). By piling the filters in a hierarchical style, respectively of which learns to trigger in reply to a different sympathetic visual input, the arrangement can progressively recognize more complex features.

$$f^{(l)} = \sigma\left(W^{(l)} * f^{(l-1)} + b^{(l)}\right) \tag{9}$$

Max Pooling Layers

Each convolutional Layer is followed by max pooling layers, which subordinate the spatial resolution of feature maps while preserving the maximum critical statistics. The network is more resilient to minor variations or misrepresentations in ulcer pictures, spatial down-sampling, which is used to increase computing efficiency and offers some translation invariance, as specified in Equation. (10). The pooling process highlights the strongest activations, which typically correlate to significant medical pointers, by procuring the maximum value in a assumed region.

$$f^{(l)} = MaxPool\left(f^{(l-1)}, k = 2\right) \tag{10}$$

Flatten Layer

The last convolutional block's 3D output is transformed into a 1D vector by the Flatten Layer. This process connects the dense layers, which carry out classification using learnt features, with the convolutional layers, which record spatial data. A simplified and abstract feature representation of the input image is represented by the flattened vector, as specified in Equations (11) and (12).

$$h = \sigma\left(w_1 \cdot flatten\left(f^{[3]}\right) + b_1\right) \tag{11}$$

$$\hat{y} = Soft\,max(\omega_2 \cdot h + b_2) \tag{12}$$

The final output $\hat{y} \in R^2$ provides class probabilities for healthy and ulcerated skin.

Dropout Layer

A technique called dropout is used to circumvent overfitting, which is particularly vital in small medical datasets where recall poses an apprehension. To strengthen the network and simplify transversely numerous paths, an arbitrary proportion (30%) of neurons in the Dropout layer are silenced throughout training, as specified in Equation (13).

$$\tilde{h} = h \circ r \tag{13}$$

$\circ$ signifies the Hadamard (element-wise) product $\tilde{h}$ The normalized feature vector is promoted to the dense Layer throughout interpretation, dropout is inactivated, and beginnings are increased to uphold reliability as denoted in Equation (14):

$$\tilde{h}_{inference} = p \cdot h \tag{14}$$

Fully Connected (Dense) Layers

High-level perception is influenced by dense layers, which are located at the network's end. The flattened feature vector is abridged into a dense method suitable for classification by the first dense Layer with ReLU activation. The model's confidence scores for the two classes of healthy and ulcerated skin are shaped by the last dense Layer, which has two output units and SoftMax activation, as denoted in Eqs. (15), (16), and (17).

$$z_1 = w_1 h + b_1 \tag{15}$$

Where h_l=max (0, z_l) (element-wise ReLU)

$$z_2 = w_2 h_1 + b_2 \tag{16}$$

$$\hat{y}_i = \frac{e^{z2i.}}{\sum_{j=1}^{2} e^{z2j}} \tag{17}$$

The model is trained using categorical cross-entropy as denoted in Equation. (18):

$$\mathcal{L}_{cE} = -\sum_{i=1}^{2} y_i \log(\hat{y}_i) \tag{18}$$

Where y_{i}_ is the one-hot encoded ground truth and $\hat{y}_i$ is the predicted softmax score for class i. The DFU-LWnet Training is outlined in Algorithm 2.

Algorithm 2: DFU-LWNet Training
Input: Dataset D = {(X_b, Y_b)}, learning rate η, batch size B, epochs E Output: Trained DFU-LWNet model
Step 1: Initialize model weights W using He initialization Step 2: For epoch = 1 to E do: a. Shuffle the dataset $\mathcal{D}$ and divide it into mini-batches of size B b. For each mini-batch of (X_b, Y_b) i. Perform forward pass to compute prediction: $\hat{Y}_b \leftarrow$ DFU_LWNet(X_b) ii. Calculate cross-entropy loss using Equation (18) iii. Backpropagate the gradients L concerning W iv. Update weights using Adam optimizer (which applies adaptive moment estimates and learning rate adjustments) Step 3: Evaluate on the validation set at the end of each epoch Step 4: Save the weights W*, which correspond to the best validation performance Return the Final trained model DFU-LWNet with weights W*

3.4. Knowledge Distillation with DFU-LWNet

Even if large models like InceptionV3 achieve excellent results on medical imaging tasks, their computational requirements make it problematic to deploy them in resource-constrained environments or in real-time. Knowledge Distillation (KD) provides a clear explanation of this complex issue by allowing a lightweight student network to benefit from the instructor's simplification without incurring the expense. This section introduces DFU-LWNet, a lightweight CNN designed for the classification of DFU. InceptionV3 bids softened estimates that unswerving learning surpasses the restraints of rigid labels, while KD is cast off to train DFU-LWNet. The accuracy of the current method, which used DFU-LWNet as the student model and InceptionV3 as the Teacher, was 96.23%.

The research utilizes this setup, but by making significant advancements in temperature scaling, loss balancing, and optimization, it is able to achieve a higher performance of 96.80%.

3.4.1. DFU-LWNet Architecture with KD

With attuned hyperparameters, the DFU-LWNet architecture used in this research closely resembles the one obtainable in existing techniques. This convolutional neural network is lightweight and designed to accurately classify medical images. The following is a description of the entire architecture as denoted in Table 3.

Table 3. DFU-LWNet architecture with KD

Layer Type	Filter/Units	Kernel Size	Activation	Output Shape
Input	-	-	-	(224, 224, 3)
Conv2D + MaxPool	32	3x3	ReLU	(112, 112, 32)
Conv2D +MaxPool	64	3x3	ReLU	(56, 56, 64)
Conv2D+MaxPool	128	3x3	ReLU	(28, 28, 128)
Flatten	-	-	-	(100352,)
Dropout	-	-	p=0.3	(100352,)
Dense	128	-	ReLU	(128,)
Dense	2	-	Softmax	(2,)

The model contains approximately 2.5 million parameters. Hierarchical feature extraction, particularly texture and edge patterns, which are vital for identifying diabetic ulcers, is supported by the convolutional layers. MaxPooling layers minimize the computational effort after reducing the spatial dimensions and maintain the pertinent characteristics. The problem of overfitting can be reduced by including the Dropout Layer. Through the final dense layer with SoftMax activation, binary classification probabilities are generated.

Binary classification probabilities are generated through a final dense layer with SoftMax activation. In contrast to the existing standard, the resolution customs optimization techniques discussed in subsequent sections aim to modify the dense layer thickness and dropout rate. The overall parameter count is ~2.5M, meaningfully slighter than InceptionV3's 28M, allowing faster inference and lower memory consumption. However, this method integrates optimized dropout and learning rate parameters, while the current method uses the same architecture.

Let $y \in \{0, 1\}^2$ be the one-hot encoded ground truth label, η_T be the SoftMax output from the teacher model, and Þ_S be the SoftMax output from the student model.

Let $x \in \mathbb{R}^{224x224x3}$ be the input image. The temperature-scaled SoftMax is expressed in Equation (19).

$$\hat{y}_i^{(\tau)} = \frac{\exp(z_i / \tau)}{\Sigma_j \exp(z_j / \tau)} \tag{19}$$

Where τ is the temperature parameter that smooths the predicted logits.

Hard Loss (Cross-Entropy) is denoted in Equation(20).

$$L_{hard} = -\Sigma_i y_i \log(\hat{y}_{S,} i) \tag{20}$$

This loss measures the distance between the one-hot labels and student predictions.

Soft Loss (KL Divergence) is denoted by Equation (21).

$$L_{soft} = \Sigma_i \hat{y}_{T,}, i^{(\tau)} \log(\hat{y}_{T,}, i^{(\tau)} / \hat{y}_{S,}, i^{(\tau)} \tag{21}$$

This loss helps the student mimic the Teacher's softened prediction distribution. Combined KD Loss is denoted by Equation (22).

$$L_{KD} = \alpha \cdot \tau^2 \cdot L_{soft} + (1 - \alpha). L_{hard} \tag{22}$$

The coefficient that stabilizes hard and soft losses is called α. The gradient magnitude is scaled by τ^2.

Wherever the two components are balanced by $\alpha \in [0, 1]$. The gradient scaling is remunerated for by τ^2.

This preparation ensures a stable influence of both hard and soft objectives, something that the existing technique did not exactly address. DFU-LWNet Student Model with KD is outlined in Algorithm 3.

Algorithm 3: DFU-LWNet Student Model with KD
Input: Dataset D = {(X_b, Y_b)}, Trained teacher T, temperature τ, balance α Output: Trained student model S
Step1: Set student model S (·) with random weights Step 2: Freeze all the layers of the pretrained teacher model T (·) Step 3: For each epoch = 1 to E, do a. For each mini-batch (X_b, Y_b) from $\mathcal{D}$ do i. Get teacher soft targets with temperature-scaled SoftMax utilizing the SoftMax Equation (19) ii. Calculate the student predictions using $\hat{y}_S \leftarrow S(X_b)$ iii. Apply the temperature-scaled SoftMax to the student outputs using Equation (19). iv. Calculate the hard loss using categorical cross-entropy using Equation (20). v. Calculate soft loss using the KL divergence using Equation (21). vi. Syndicate losses using Equation (22). vii. Bring up-to-date student model S via backpropagation using L_{KD} Step 4: Return the trained student model S.

3.5. DFU-LWNet with Knowledge Distillation and Binary Whale Optimization Algorithm (BWOA)

The Binary Whale Optimization Algorithm (BWOA) is a metaheuristic search technique that draws inspiration from the communal hunting habits of humpback whales. The original Whale Optimization Algorithm (WOA) was modified to generate BWOA, which is particularly suitable for binary search spaces, such as feature selection and hyperparameter tuning [7]. This method utilizes BWOA to tune the hyperparameters of the proposed student model, DFU-LWNet, in a knowledge distillation framework for DFU classification.

The addition of BWOA to the proposed pipeline was motivated by two goals: to increase the lightweight model's classification accuracy through hyperparameter tweaking, and to automate and validate the procedure of detecting the best configurations, thereby reducing manual labour and guesswork [37]. The combination of metaheuristic optimization and Knowledge Distillation (KD) in DFU-LWNet proposals presents a collaborative method that meets the efficiency and accuracy requirements of DFU classification.

KD handovers softened output distributions from a deep, pretrained teacher model to the student model, swelling its representative capability. Here, this method utilizes the Binary Whale Optimization Algorithm (BWOA) for automatic hyperparameter tuning to suggestively enhance DFU-LWNet's performance, thereby circumventing the disadvantages of traditional search.

The reasoning, development, and application of KD in conjunction with BWOA are presented in this section. The end product is a lightweight, efficient network that may preserve deployability while reaching classification performance comparable to sophisticated designs. In KD, the student network learns from both the soft labels Þ_T^((T)) produced by the teacher network at a higher temperature, as well as the hard labels y ∈ {0, 1} ². The two goals are balanced in the final KD loss function, as expressed in Equation (23).

$$L_{CE} = \propto . T^2 . KL\left(\hat{y}_T^{(T)} | \hat{y}_S^{(T)}\right) + (1 - \alpha). CE(y, \hat{y}_S) \quad (23)$$

Humpback whales search for food using a superior technique known as bubble-net feeding. It includes two Primary techniques: the whales socialize everywhere their prey is in a spiral or weakening circle after approximating its location. Whales engage in exploitation and exploration as an alternative to lengthening their search in the vicinity of potential solutions and altering their course to steer clear of these local optima. DFU-LWNet's discrete hyperparameter space is examined using BWOA. Among the hyperparameters that were optimized are:

- Dropout rate: {0.2, 0.4}
- Dense layer units: {64, 128, 256}
- Learning rate: {1e-4, 5e-4, 1e-3}

Each candidate solution (agent) in the whale population is a binary string encoding one choice per hyperparameter, as denoted in Equation (24).

$$A = 2a. r - a, C = 2. r, a = 2 - \frac{2t}{t} D = |C - G^* - X_i^t| and X_i^{t+1} = G^* - A. D \quad (24)$$

Binary adaptation uses a sigmoid transfer function as denoted in Equation (25).

$$S(X) = \frac{1}{1+e^{-x}} => X_i = \begin{cases} 1 \text{ if } S(X) > r \\ 0 \text{ otherwise} \end{cases} \quad (25)$$

The DFU-LWNet with KD and BWOA is outlined in Algorithm 4.

Algorithm 4: DFU-LWNet with KD and BWOA
Input: Dataset D, teacher T, search space H, KD params (τ, α), BWOA params Output: Optimized student model S*
Step 1: Use random binary vectors from H2 to initialize the BWOA population. Step 2: For every whale (agent) in the population: Decode the hyperparameters (LR, dense, dropout). b. Use the current parameters to train DFU-LWNet with KD. c. Assess the accuracy of validation and update fitness Step 3: Use the BWOA update rules to update whale positions.

Step 4: Continue until convergence is achieved or the maximum number of iterations is reached.

Step 5: Retrain the final DFU-LWNet after choosing the optimal hyperparameter set.

This method closes the performance gap between deep and lightweight CNNs by combining BWOA and KD together. DFU-LWNet becomes a more powerful diagnostic tool with the help of InceptionV3 and BWOA refinement in real-time and with higher accuracy [32]. This method creates a scalable framework for the exact AI healthcare systems.

4. Results and Discussion

4.1. Evaluation Metrics

This research method uses a range of statistical and computational measures to evaluate the performance of the models in DFU detection [18].

The metrics were chosen to provide a clear view of diagnostic consistency, and they are mostly used in medical imaging. The statistical measures are model size, parameter count, inference time, recall (REC), specificity (SPEC), accuracy (ACC), precision (PRE), and F1-score.

4.1.1. Accuracy (ACC)

Accuracy is the proportion of the correctly identified cases and includes the true positive and true negative in the total number of samples. In a medical dataset, if there is a class imbalance, which is a common issue in medical datasets, the final accuracy will provide a misleading impression of the whole performance of the system.

4.1.2. Recall (TPR) or Sensitivity

The model is correctly identified, and recall is the proportion of true positives. This is very crucial in medical datasets, particularly for DFU diagnosis, as it is a missed diagnosis and will lead to severe consequences.

4.1.3. Specificity (TNR)

Specificity ensures that healthy and ulcerated images are not mistakenly identified, and this measure is used to determine the actual negative rate and help avoid false alarms.

4.1.4. Inference Time

This is the number of milliseconds (ms) needed to classify one image. Low inference times are crucial for real-time applications, particularly in embedded and mobile medical devices.

4.1.5. Parameter Count and Model Size

Parameter count is used to measure the storage footprint, usually in MB, and also measures the quantity of trainable parameters. Smaller models find difficulty with deployment in edge devices with constrained memory and storage, smaller than the models that are crucial. The InceptionV3 teacher model provides higher classification accuracy and contains approximately 28 million parameters. It requires 90 MB of storage, and this cannot be done for real-time deployment.

The proposed model, DFU-LWNet, is designed to be more compact. It contains approximately 2.1 million trainable parameters and comprises a model size of 8.4 MB. Knowledge distillation improves DFU-LWNet learning capabilities without increasing its model size, and the architectural structure is not changed.

Binary whale optimization is used for optimization, and hyperparameter tuning is done. There is a moderate increase in the parameter count with 2.3 million parameters and a model size of 9.2 MB. This is suitable for edge deployment and provides improved classification performance.

4.2. Implementation Details and Hyperparameter Selection

The implementation of the model is done in Python using TensorFlow 2.9 and the Keras Library. Uses NVIDIA Tesla GPU and 16 GB of RAM in Google Colab. Early stopping, model checkpointing, and real-time data augmentation are done to ensure industry best practices and adhere to them.

The dataset used in this approach is a publicly available dataset that is separated into 2 categories: Healthy and Ulcerated skin. Each image in the dataset was resized to 224 × 224 pixels. The preprocessing pipeline was normalized to the [-1, 1] range by following illumination normalization and using Contrast-Limited Adaptive Histogram Equalization (CLAHE).

4.3. Data Partitioning

- 80% of the training set
- 10% is the validation set.
- 10% is the test set.
- 32 is the batch size.
- Optimizer-Adam.
- Epochs- Up to 30 epochs with early stopping after five consecutive non-improving epochs.
- The Learning Rate was initially set and dynamically adjusted during tuning.

4.3.1. Hyperparameters in BWOA

- [0.2,0.4] are the rates of dropout.
- [64, 128, 256] are the units of the Dense Layer.
- [1e-4, 5e-4, 1e-3] are the learning rates.

The Binary Whale Optimization Algorithm was run with population sizes of five and ten, with ten iterations in order to balance search exploration and convergence duration. This is done in contrast to the existing method that employed a set configuration of learning rate (1e-4), dense units (128), and dropout (0.3).

4.4. Teacher Model Performance: InceptionV3

The instructor model for knowledge distillation is InceptionV3.InceptionV3 has a comprehensive architecture and has shown efficiency in medical imaging tasks. This method combines auxiliary classifiers, factorization, and convolutions to enhance feature extraction at multiple level. DFU-LWNet is a lightweight student model for edge development in an environment that has fewer resources.

DFU-LWNet was trained using traditional supervised learning on a ground truth table in its baseline configuration, and this is done using knowledge distillation (KD). There are three convolutional blocks, and each is followed by batch normalization, ReLU activation, and max-pooling layers, as well as by flattening and a fully connected classification head.

This model demonstrates good baseline results with an average inference time of 22 ms per image and provides an accuracy of 94.12% with 2.1 million parameters. This method has a substantially lower computing cost that meets clinical diagnostic requirements and gives slightly lower performance than InceptionV3. DFU-LWNet is suitable for embedded applications, and this has the capacity to overpredict ulcers because of its limited capacity.

4.5. DFU-LWNet with Knowledge Distillation (KD)

DFU-LWNet's Classification improved by incorporating KD into its training pipeline. The student model was trained to replicate the probabilistic behavior of the teacher model by using the soft logits instead of completely relying on binary labels. This enables a richer training signal and captures similarities between classes. DFU-LWNet performance increased with KD, and accuracy rises to 96.23% and the F1-Score increases to 96.32%.

The same configuration is repeated as in the existing methodology and confirms the generalizability of the KD technique. There is a slight difference in inference time, which is 1 ms, and this can be neglected when comparing to the diagnostic advantages.

4.6. DFU-LWNet with KD and Binary Whale Optimization Algorithm (BWOA)

The limits of the KD-based DFU-LWNet were tested by automatically adjusting three key hyperparameters -learning rate, dense layer size, and dropout rate - using BWOA. The whale population changed over ten cycles, steadily increasing its fitness score (1 - validation accuracy).

The optimal configuration discovered was:

- Dropout: 0.2
- Dense Units: 256
- Rate of Learning: 5e-4

This configuration yielded the best-performing model among all evaluated versions, outperforming both manually and grid-search-tuned counterparts. In comparison to the fixed KD configuration in the current method, BWOA improved accuracy by 0.57% and F1-score by 0.6% while maintaining computing performance. The slight increase in parameter count (from 2.1 million to 2.3 million) is worthwhile for the improvement in clinical performance.

This result demonstrates that BWOA can effectively identify configurations that are close to optimal through binary search.

4.7. Ablation Study

To assess the distinct effects of Knowledge Distillation (KD) and the Binary Whale Optimization Algorithm (BWOA) on the final model's performance, this research conducted an ablation study. DFU-LWNet was systematically evaluated in four configurations. The goal of the experiment is to determine the optimal configuration due to its computational efficiency and performance, and to investigate the impact of each element on the final classification metrics.

The results of the experiment show how the Knowledge distillation method transfers semantic information from the instructor model and increases its performance. There is a slight increase, which is offered by grid search, and that signifies the limited capacity to investigate its optimal configurations. BWOA provides the best performance improvement at a very minimal computational cost.

The enhanced model is the same as the original model but gives a 10x reduction in model size and over 3x faster inference. This makes it more ideal for mobile health applications, mainly in rural areas or areas that have poor resources. The performance difference between KD-fixed and KD-BWOA appears negligible when tested on large screening datasets but leads to an important reduction in misclassified instances.

4.8. Inference Time Analysis

Inference time is an important metric in real-time diagnostic systems. If there is a long inference time, it reduces user confidence in the mobile application and delays the clinical diagnosis.

Table 4. Inference time analysis

Models	Inference Time Analysis
InceptionV3	89 ms
DFU-LWNet (Baseline)	22 ms
DFU-LWNet - KD	23 ms
DFU-LWNet-KD with BWOA	24 ms

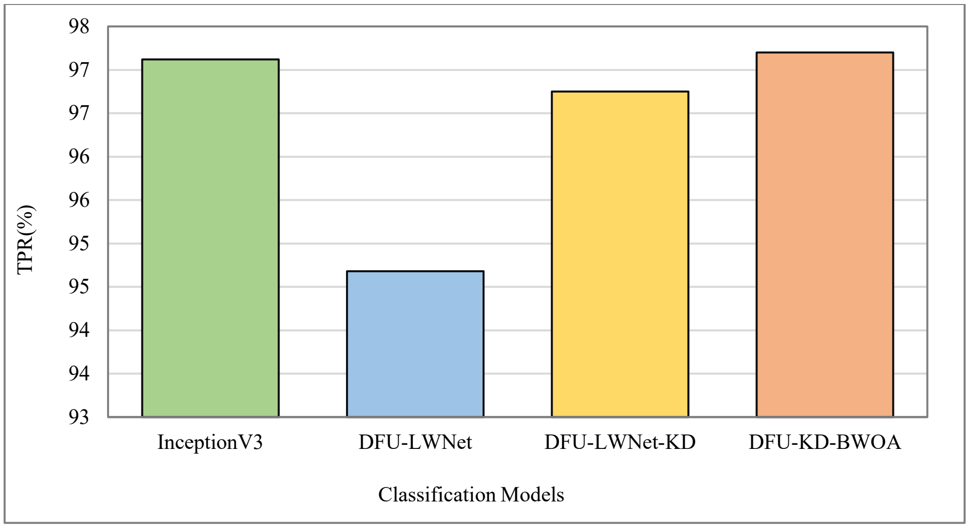

Fig. 2 TPR analysis of DFU on different deep learning models

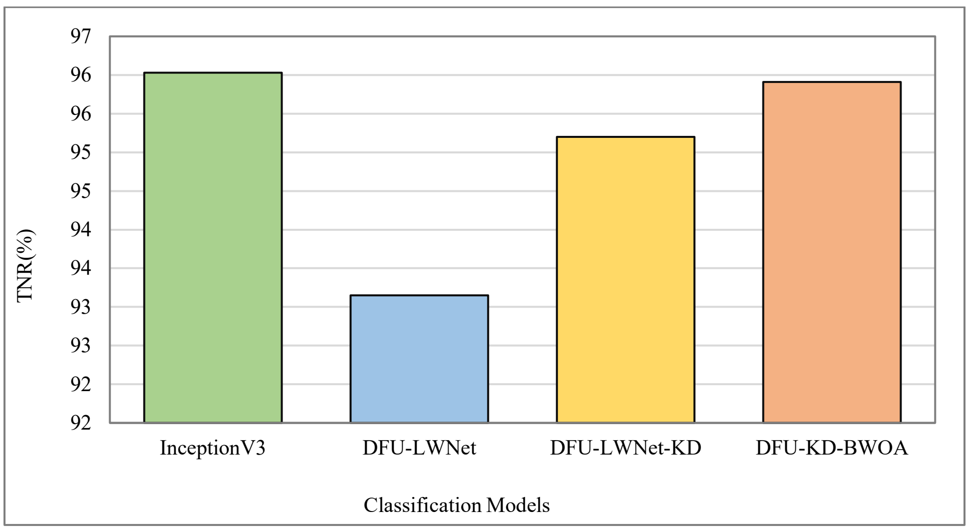

Fig. 3 TPR analysis of DFU on different deep learning models

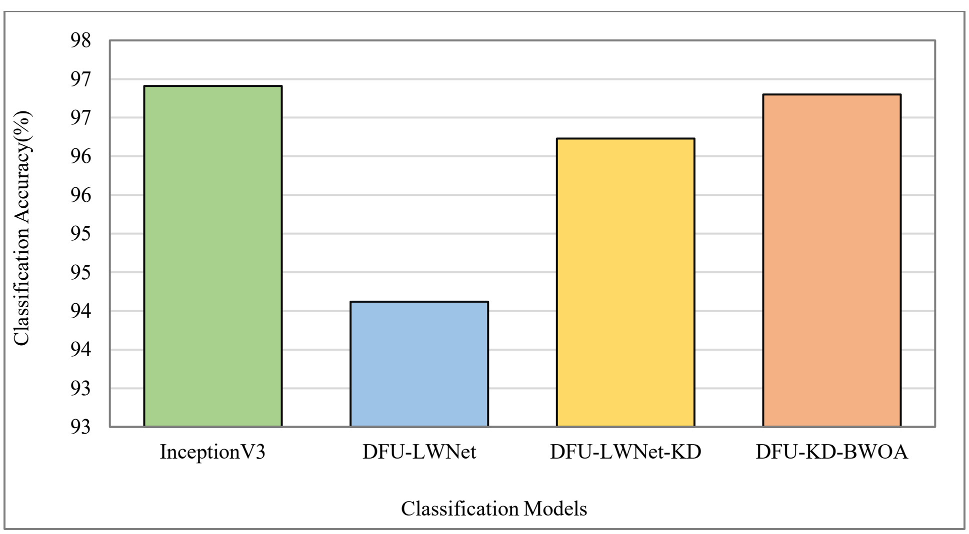

Fig. 4 Classification accuracy analysis of DFU on different deep learning models

The inference benchmarks between the models show a notable difference in latency and computational load. The proposed model takes four times as long to deduce compared to the teacher model. This shows the usefulness of the compact model that can be enhanced by intelligent optimization and knowledge distillation. BWOA shows a slight increase in the number of parameters and inference time, and stays within the acceptable bounds for real-time applications. The method KD with the BWOA model demonstrates improved stability, which is reliable for embedded deployment with the lowest variation in the inference time. The proposed method, DFU-LWNet-KD with BWOA, shows the best balance between computational cost, performance, and reliability. InceptionV3's high accuracy disproves its computational cost and unsuitability for real-time and in deployment with fewer resources.

5. Conclusion

This research method develops a robust and lightweight deep learning architecture for classifying DFU images using a novel combination of Knowledge distillation and the Binary Whale Optimization Algorithm (BWOA). The recommended student model, DFU-LWNet, was designed for efficient implementation in an environment with fewer resources. A less resource-intensive environment can be embedded in a mobile medical diagnostic device. The outcome provides implementation of the approach in a feasible way. The teacher model, with an accuracy of 96.91% performs as the performing teacher model. The teacher model is difficult to implement because of its numerous parameters and inference time, mainly in environments with fewer resources. DFU-LWNet shows a reasonable balance between accuracy and efficiency in its basic setup. DFU-LWNet accuracy increases by 96.23% after doing KD training and demonstrates the efficiency of teacher semantic feature knowledge transfer. The important change occurred after the addition of BWOA, which automatically increases the student model hyperparameters. DFU-LWNet achieves 96.80% accuracy with 2.3 million parameters and 24 ms inference time. This kind of integration demonstrates the dual benefit of the method, which has high accuracy and efficiency. This methodology ensures that this can be applied to medical image classification issues beyond DFU. This method outperforms the existing method and the baseline system described in the existing methodology by replacing empirical tuning with a good optimization strategy. The integration of knowledge distillation and metaheuristic optimization into a lightweight CNN architecture is an innovative approach for medical image diagnostics. This kind of platform pushes the limits of performance and makes it easier to use a small deep learning model in practice. In the future, this research aims to extend to multi-class DFU phases and integrate temporal data from longitudinal case studies, and try to enhance its prediction modeling. This work aims to provide practical implications for real-world DFU monitoring and deliver an accurate, lightweight, and deployable model suitable for clinical settings with low resources. The framework can be extended in the future to multi-class stages and higher clinical followers. The future work also aims at the integration of longitudinal data and validates the system across the globe with diverse patients, and enhances generalizability.

References

[1] Moi Hoon Yap et al., "Deep Learning in Diabetic Foot Ulcers Detection: A Comprehensive Evaluation," *Computers in Biology and Medicine*, vol. 135, pp. 1-17, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] Fitri Arnia et al., "Towards Accurate Diabetic Foot Ulcer Image Classification: Leveraging CNN Pre-Trained Features and Extreme Learning Machine," *Smart Health*, vol. 33, pp. 1-12, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Kamran Amjad et al., "A Novel Lightweight Deep Learning Framework with Knowledge Distillation for Efficient Diabetic Foot Ulcer Detection," *Applied Soft Computing*, vol. 167, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Jinhai Fang et al., "Anomaly Detection of Diabetes Data based on Hierarchical Clustering and CNN," *Procedia Computer Science*, vol. 199, pp. 71-78, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] André Kharitonov et al., "Literature Survey on Combining Machine Learning and Metaheuristics for Decision-Making," *Procedia Computer Science*, vol. 253, pp. 199-208, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[6] Juan Pablo Mesa et al., "Machine-Learning Component for Multi-Start Metaheuristics to Solve the Capacitated Vehicle Routing Problem," *Applied Soft Computing*, vol. 173, pp. 1-32, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[7] Dil Bahar, Akshay Dvivedi, and Pradeep Kumar, "Optimizing the Quality Characteristics of Glass Composite Vias for RF-MEMS using Central Composite Design, Metaheuristics, and Bayesian Regularization-Based Machine Learning," *Measurement*, vol. 243, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[8] Sulaiman Afolabi et al., "Informatics and Health Predicting Diabetes using Supervised Machine Learning Algorithms on E-Health Records," *Informatics and Health*, vol. 2, no. 1, pp. 9-16, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[9] Averina Nurdin et al., "Using Machine Learning for the Prediction of Diabetes with Emphasis on Blood Content," *Procedia Computer Science*, vol. 227, pp. 990-1001, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Durga Parkhi et al., "Prediction of Postpartum Prediabetes by Machine Learning Methods in Women with Gestational Diabetes Mellitus," *iScience*, vol. 26, no. 10, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Arinze Nkemdirim Okere et al., "Evaluation of Factors Predicting Transition from Prediabetes to Diabetes among Patients Residing in Underserved Communities in the United States - A Machine Learning Approach," *Computers in Biology and Medicine*, vol. 187, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[12] Yitayeh Belsti et al., "Comparison of Machine Learning and Conventional Logistic Regression-based Prediction Models for Gestational Diabetes in an Ethnically Diverse Population; The Monash GDM Machine Learning Model," *International Journal of Medical Informatics*, vol. 179, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Fang Zhou et al., "A Stepwise Prediction and Interpretation of Gestational Diabetes Mellitus: Foster the Practical Application of Machine Learning in Clinical Decision," *Heliyon*, vol. 10, no. 12, pp. 1-11, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Liangjun Jiang et al., "Diabetes Risk Prediction Model based on Community Follow-Up Data using Machine Learning," *Preventive Medicine Reports*, vol. 35, pp. 1-16, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Muhammad Exell Febrian et al., "Diabetes Prediction using Supervised Machine Learning," *Procedia Computer Science*, vol. 216, pp. 21-30, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Ketan Dhatariya, and Zulfiqarali G. Abbas, "Estimated Costs of Treating Two Standardised Diabetes-Related Foot Ulcers of Different Severity - A Comparison of 7 Global Regions," *Diabetes Research and Clinical Practice*, vol. 221, pp. 1-7, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[17] Puneeth N. Thotad, Geeta R. Bharamagoudar, and Basavaraj S. Anami, "Diabetic Foot Ulcer Detection using Deep Learning Approaches," *Sensors International*, vol. 4, pp. 1-9, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Shuvo Biswas et al., "DFU_Multinet: A Deep Neural Network Approach for Detecting Diabetic Foot Ulcers through Multi-Scale Feature Fusion using the DFU Dataset," *Intelligence-Based Medicine*, vol. 8, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Malik Adnan et al., "IDC Pressure Sensors Enabled Smart Footwear System for in Vitro Detection and Monitoring of Diabetic Foot Ulcer," *Sensors and Actuators A: Physical*, vol. 388, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[20] Mrinal Kanti Dhar et al., "Fusegnet: A Deep Convolutional Neural Network for Foot Ulcer Segmentation," *Biomedical Signal Processing and Control*, vol. 92, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] Marizuana Mat Daud et al., *Revolutionizing Diabetic Foot Ulcer Treatment Prediction: Harnessing the Power of Artificial Intelligence and Transfer Learning*, Uncertainty in Computational Intelligence-Based Decision Making: Advanced Studies in Complex Systems, pp. 55-63, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[22] Haidong Ye, and Fang Yao, "Deep Learning-Enhanced MRI for Diabetic Foot Tarsal Bone Lesions and Insulin Injection Behavior Analysis," *Journal of Radiation Research and Applied Sciences*, vol. 18, no. 2, pp. 1-10, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[23] N. Christy Evangeline, and S. Srinivasan, "Deep Neural Net for Identification of Neuropathic Foot in Subjects with Type 2 Diabetes Mellitus using Plantar Foot Thermographic Images," *Biomed Signal Process Control*, vol. 96, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] Andrés Anaya-Isaza, and Martha Zequera-Diaz, "Fourier Transform-Based Data Augmentation in Deep Learning for Diabetic Foot Thermograph Classification," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 2, pp. 437-452, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[25] Lit Zhi Chee et al., "Gait Acceleration-Based Diabetes Detection Using Hybrid Deep Learning," *Biomedical Signal Processing and Control*, vol. 92, pp. 1-8, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[26] Farman Ali et al., "IP-GCN: A Deep Learning Model for Prediction of Insulin using Graph Convolutional Network for Diabetes Drug Design," *Journal of Computational Science*, vol. 81, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[27] Keona Pang, "A Comparative Study of Explainable Machine Learning Models with Shapley Values for Diabetes Prediction," *Healthcare Analytics*, vol. 7, pp. 1-11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[28] Teagan Weatherall, et al., "The Impact of Machine Learning on the Prediction of Diabetic Foot Ulcers - A Systematic Review," *Journal of Tissue Viability*, vol. 33, no. 4, pp. 853-863, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[29] Shuvo Biswas et al., "XAI-Fusionnet: Diabetic Foot Ulcer Detection based on Multi-Scale Feature Fusion with Explainable Artificial Intelligence," *Heliyon*, vol. 10, no. 10, pp. 1-23, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[30] Jenny Elizabeth Price et al., "Machine Learning Algorithms Mimicking Specialists Decision Making on Initial Treatment for People with Type 2 Diabetes Mellitus in Japan Diabetes Data Management Study (JDDM76)," *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, vol. 18, no. 11-12, pp. 1-6, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[31] Lingga Aksara Putra et al., "State Estimation of a Biogas Plant based on Spectral Analysis using a Combination of Machine Learning and Metaheuristic Algorithms," *Applied Energy*, vol. 377, pp. 1-15, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[32] Sagarika Mohanty, Bibhudatta Sahoo, and Subham Sai Behera, "An Assessment of Nature-Inspired Metaheuristic Algorithms for Resilient Controller Placement in Software-Defined Networks," *Decision Analytics Journal*, vol. 12, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[33] Ramandeep Saha, and Somnath Pal, "A Hybrid Metaheuristic Algorithm using Elitist Chemical Reaction Optimization and Learning from Knowledge Assimilation for Improving Rule-based Classification Models," *Procedia Computer Science*, vol. 235, pp. 701-712, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[34] Ruben Silva-Tinoco et al., "Improving Foot Ulcer Risk Assessment and Identifying Associated Factors: Results of an Initiative Enhancing Diabetes Care in Primary Settings," *Diabetes Epidemiology and Management*, vol. 14, pp. 1-8, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[35] Robert Fitridge et al., "The Intersocietal IWGDF, ESVS, SVS Guidelines on Peripheral Artery Disease in People with Diabetes Mellitus and a Foot Ulcer," *Journal of Vascular Surgery*, vol. 78, no. 5, pp. 1101-1131, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[36] Gyeongho Kim et al., "Developing a Data-Driven System for Grinding Process Parameter Optimization using Machine Learning and Metaheuristic Algorithms," *CIRP Journal of Manufacturing Science and Technology*, vol. 51, pp. 20-35, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[37] Jaesun Park, Jeahoon Cho, and Kyung-Young Jung, "Nature-Inspired Metaheuristic Optimization Algorithms for FDTD Dispersion Modeling," *AEU - International Journal of Electronics and Communications*, vol. 187, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]