Journal of Engineering Science and Technology
Vol. 20, No. 2 (2025) 577 - 589
© School of Engineering, Taylor's University

MOBILEDFU: CLASSIFICATION OF DIABETIC FOOT ULCER INFECTION ON THE EDGE

AUDREY K. C. HUONG*,
JALALUDDIN MUHAMMAD SUKKI, XAVIER T. K. NGU

Faculty of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Johor, Malaysia
*Corresponding Author: audrey@uthm.edu.my

Abstract

The standard diagnosis of diabetic foot ulcer (DFU) infection is limited to scheduled visits, and late treatment of this disease can result in further complications and limb amputation. This study proposed an automatic framework for rapidly detecting DFU infection status on the edge. The design of this framework used Google Colab and Android Studio IDE platforms to train and optimize the employed models to run on a mobile device using the developed Android app. DFU image classification results are compared for pretrained MobileNetV2 and EfficientNetB3, and a custom-developed stack convolutional model optimized for the two-class classification problem. The testing results showed that MobileNetV2 and EfficientNetB3 systems recognized DFU infection status with a classification accuracy of 0.74 and 0.8, respectively, whereas the stack model performed comparatively inferior. Most misclassified images suffered from severe motion blurs and loss of finer details. This paper recognized MobileNetV2 as the benchmark model for further exploration in this study area. The future of this work includes adopting a more effective weight updating process and incorporating preprocessing steps to enhance image quality and remove background prior to the prediction. This proposed strategy could improve patient care quality in DFU management by supporting real-time clinical decision-making in at-home or remote care settings, reducing morbidity and mortality.

Keywords: Android, DFU, Edge, Infection, MobileNetV2.

1. Introduction

With the advancement of artificial intelligence (AI) applications in all areas of life, this technology has significantly transformed how humans work, interact, live, and carry out their everyday tasks. The ability of this technology to learn from large datasets or past computations eases the decision-making process, allowing efficient knowledge and resource management while profoundly improving human well-being and living qualities. The application of AI is vast and spans from heavy industry and manufacturing to AI-powered technology and services [1]. Its application in healthcare is becoming increasingly prevalent to improve delivery and promote best practices. The integration of AI with imaging technologies has been proposed to diagnose diseases, such as infections, malignancies, and neurological disorders [2-4].

Besides, combining AI with telemedicine for home-based care support allows the diagnosis to be made immediately using an edge device, i.e., a smartphone, thus improving clinical outcomes and patient experience. Previous efforts that adopted human-cantered AI for mobile diagnostics and treatment recommendations include skin cancer detection using a convolutional neural network (CNN)-enabled mobile app in [5], Internet of Things (IoT) technologies proposed for improved self-management of chronic diseases during the COVID-19 pandemic in [6-8], and mobile system for rehabilitation and recovery applications [9].

Diabetic foot ulcer (DFU) is one of the non-communicable chronic diseases suffered by a third of type II diabetic patients worldwide [10]. This complication is caused by poor circulation of oxygen-carrying blood flowing to the lower extremities due to the high glucose microenvironment, thus causing vascular endothelial dysfunctions [10]. DFU can be classified into different disease states: negative or positive infection, ischemia (inadequate blood supply), or a combination of both [11]. Lower limb ischemia is an important prognosis factor that leads to DFU development, delayed healing of the wound, and further patient morbidity. This pathological condition causes the narrowing of arteries, resulting in soft tissue infections. Although these clinical problems are reversible, they are chronic and recurrent. If the infected wound progresses to an advanced stage, the infection could be spread rapidly to the contiguous subcutaneous skin and bone, causing osteomyelitis [12], which can be life-threatening. Therefore, limb amputation is often necessary to halt the progression of the disease and prevent mortality.

Nonetheless, many patients do not seek proper medical attention until the condition worsens, largely due to their negligence or lack of appropriate knowledge about the disease. This problem is further exacerbated if the disease has been wrongly diagnosed, which can lead to the improper treatment of the patient, causing further complications. Unlike DFU infection, ischemic wounds are easier to recognize based on necrosis and gangrene appearance [13]. The conventional procedure of DFU infection detection is based on visual examination using magnification tools and clinical scoring by an experienced orthopaedic. The affected soft tissue specimen would also be collected from biopsy, swab culture, needle aspiration, and probe to bone (PTB) [13, 14] to confirm the severity of the infection.

The abovementioned issues call for an early and prompt appropriate treatment in DFU patient care and strategy to improve clinical outcomes and living quality [14]. Since most diagnoses and treatments can only be performed in clinical settings and during the scheduled clinical visits, which is the primary point of contact between the public and healthcare providers, quality diabetic patient care should be extended into home settings. This is possible following the advancement of AI and IoT technologies that warrant the integration of millions of devices and large data processing capabilities, thus strengthening the digitalization of healthcare. This infrastructure supports better health management procedures and more efficient quality care [4].

Among the earlier studies in the diabetes self-management domain worth highlighting are Joachim et al. [15], who demonstrated using a cross-platform mobile application and a web-based application for personalized fitness and nutrition recommendations for diabetic patients. Others include Tripathi et al. [16], who proposed using an IoT-based spatiotemporal CNN system for continuous blood glucose monitoring to improve diabetes management, while another study of Shen et al. [17] recommended a smartphone app that supports gestational diabetes diagnosis using fasting glucose data in pregnant women. Although AI-based diabetic care and research are well underway, there remains work to be done for better management and prevention of deterioration of foot ulcers, which is the most dreaded complication of diabetes, using mobile applications.

Based on the urgency of the need for more efficient management of DFU infection described above, this study proposes an innovative use of AI in developing a home-assisted healthcare system for mobile and rapid DFU infection detection. This strategy facilitates rapid and reliable diagnosis based on the acquired image, complementing the standard screening procedures, and enabling patients to seek prompt clinical confirmation and treatment from healthcare facilities in cases of suspected positive infection.

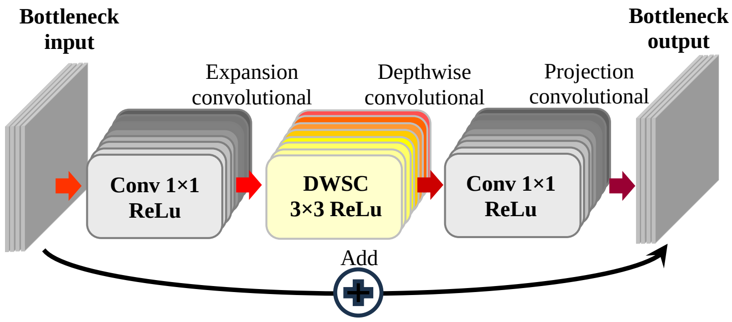
2. Methods

2.1. Convolutional neural networks and model training

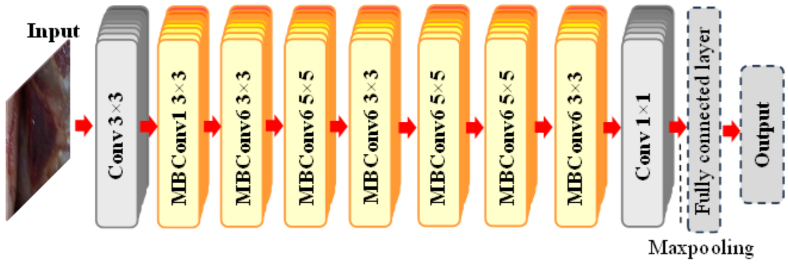
The primary objective of this paper is to propose a lightweight system for DFU infection status classification, which can be deployed on an Android mobile device for inferencing. For this purpose, this work used open-source Google Colab for the simulation. This platform supports importing Keras and TensorFlow (TF) API libraries for machine learning applications and edge model optimization and compilation using Google's free cloud service. All experiments were implemented with run time Python 3 on an NVIDIA T4 GPU and 78 GB of RAM.

For the sake of experiments, pre-trained convolutional (conv) MobileNetV2 and EfficientNetB3, shown in Figs. 1(a) and (b), are used in this study. Their weights loaded from the repository of the TensorFlow hub are used as a feature extractor. MobileNet is a lightweight, depth-wise separable convolution (DWSC) model consisting mainly of inverted residuals and linear bottleneck blocks, whilst EfficientNet is a scaling method that uses inverted residual blocks (MBConv) for efficient feature extraction.

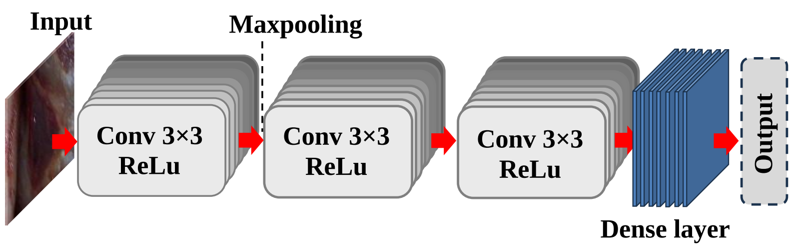
The neuron weights of these models were pre-trained on the ImageNet dataset (ILSVRC-2012-CLS) and were shown to outperform their predecessors, e.g., MobileNetV1, and other traditional convolutional models, such as ResNet and Inception, VGG, in terms of performance and computational cost [18, 19]. These models are used with transfer learning to classify DFU infection images. This work also proposes a custom-developed CNN formed by linearly stacking three convolutional layers shown in Fig. 1(c) for the two classes of DFU infection classification with end-to-end training.



(a) MobileNetV2



(b) EfficientNetB3



(c) Stack convolutional model

Fig. 1. Convolutional neural network architecture of (a) MobileNetV2, (b) EfficientNetB3 and (c) Stack convolutional model.

The dataset used for models' training was selected from the Diabetic Foot Ulcers Grand Challenge (DFUC-21) dataset (Part B) from [11] containing 15,760 DFU images collected from Lancashire Teaching Hospital. This dataset comprises 5,890 and 9,870 images annotated and grouped according to their infection and ischemia status, respectively, by a podiatrist and a DFU consultant physician. This study considered only the infection problem; thus, images of the positive and negative (non-infection) classes are used in the training and testing. Examples of these images and their ground truth label are shown in Fig. 2.

Fig. 2. DFU infection images and their ground truth label.

Each image is 256 × 256 pixels in size, which is resized to 224 × 224 pixels and 299×299 pixels, respectively, according to the input size requirement of MobileNetV2 and EfficientNetB3. The stack convolutional model in Fig. 1(c) has the same input size as the EfficientNetB3. The balanced positive and negative infection classes of 2,945 images (each class) were randomly split into training, validation, and testing datasets using a split ratio of 85 %, 10 %, and 5 %. Considerable effort (by manual cross-checking) has been taken to avoid data contamination between training and testing sets. The testing set was set aside for final system validation implemented on a mobile Android device. The training set is fed into the input layer of the network, which is connected to the feature extractor in Fig. 1, followed by a dense output layer of two classes. These layers were stacked, and these models were built using Keras sequential API. Figure 3 shows the overall workflow in developing the deep learning model for edge device deployment. The resized DFU images from the preprocessing steps are used for network training.

The 20-epoch model training was performed using SGDM as an optimizer with a mini-batch size of 64 and an initial learning rate of 0.005 to fine-tune the weight of the network neurons for the problem. The trained model is then converted from TensorFlow to a TensorFlow lite file format (.tflite), which is compatible with Android devices.

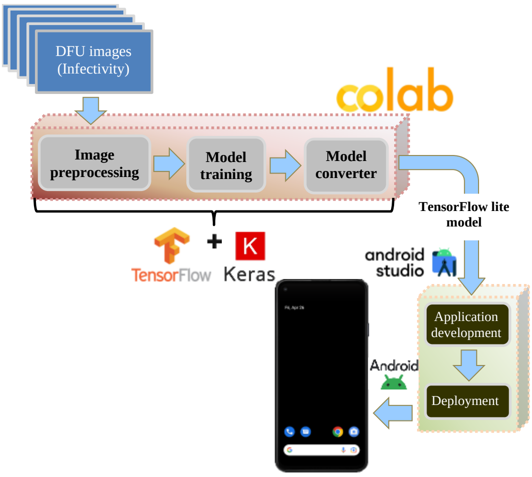


Fig. 3. Machine learning pipeline and mobile app development flow.

2.2. Edge deployment and prediction

The mobile application (app) for real-time diagnosis of DFU infection was designed and built using Android Studio Flamingo (2022.2.1) with Java programming. The functional parts of the app and the designed user interface (UI) layout for image acquisition, selection, and classification are shown in Fig. 4. The classification is performed using the generated *.tflite* file imported into the platform. Users can select the existing images from the gallery or capture them through the camera. The chosen image would be resized to 224×224 or 299×299 for conformity with the input layer of the models in Fig. 1 before being displayed on the viewing panel for visualization. The predicted classification result would be displayed on the UI screen in real-time. The custom-built app was deployed on Xiaomi Redmi Note 4 with Android 7.0 operating system for testing and demonstration. This device has 3GB RAM and an Octa-core CPU with Max 2GHz for the classification of DFU images without cloud support.

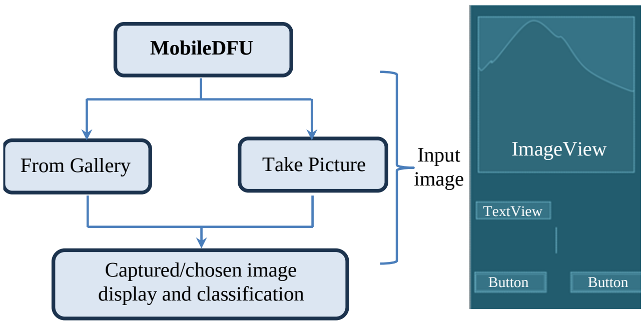


Fig. 4. The functional parts of the developed application (left) and the designed user interface (right).

2.3. Performance evaluation metrics

The performance of the classification models in their ability to identify the wound infection status is evaluated based on the prediction accuracy (*ACC*), precision (*PREC*), sensitivity (*SENS*), specificity (*SPEC*), and F1 score (*F*1) shown in Eqs. (1) – (5).

$$ACC = \frac{(TP + TN)}{(FN + FP + TP + TN)} \quad (1)$$

$$SENS = \frac{TP}{(TP + FN)} \quad (2)$$

$$SPEC = \frac{TN}{(TN + FP)} \quad (3)$$

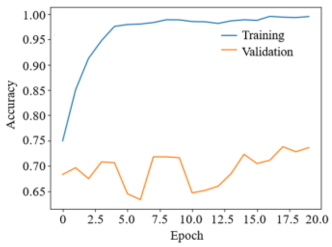
$$PREC = \frac{TP}{(TP + FP)} \quad (4)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (5)$$

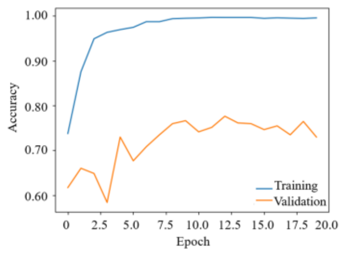
Where True Positive (*TP*) is the correct prediction of an infected wound image, True Negative (*TN*) is an accurate prediction of the non-infected DFU. False Negative (*FN*) is the probability of an infected wound (i.e., positive result) being misclassified as normal/non-infected (i.e., negative result), and False Positive (*FP*) is the opposite concept of *FN*. Thus, the *ACC* metric in Eq. (1) describes the ratio of the number of samples correctly classified by the system to the total number of samples. *SENS* or recall is the ratio of *TP* (positive infection) predictions to the total number of actual positive cases, whereas *SPEC* is the ability of the model to classify *TN* cases among all negative cases correctly. *PREC* in Eq. (4) measures the system's ability to predict *TP* cases among all positive predicted cases. *F*1 in Eq. (5) is computed as twice the product of *PREC* and *SENS* divided by their sum.

3. Results

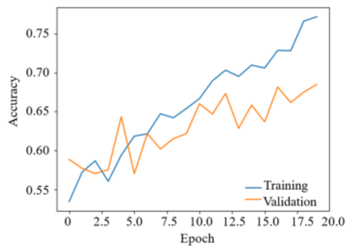
The training process in this study uses 20 epochs to optimize the network performance. Figure 5 shows the training and validation accuracies for each iteration of the epoch. The figure shows that the model training performance improves with epoch number. The training accuracy of pretrained networks (panels a and b of Fig. 5) reached a plateau, accompanied by a slow increase in the validation accuracy after the fifth epoch, suggesting the possibility of models' overfitting beyond this epoch.



(a) MobileNetV2



(b) EfficientNetB3



(c) Stack convolutional model

Fig. 5. Model training and validation progress.

The training process took approximately 20 minutes for MobileNetV2 and the stack model, while EfficientNetB3 required 45 minutes to execute. The testing confusion matrix of these models (in *tflite* format on an edge device) evaluated using testing images is shown in Fig. 6. The performance measures in Eqs. (1)-(5) calculated based on the results in the figure are tabulated in Table 1.

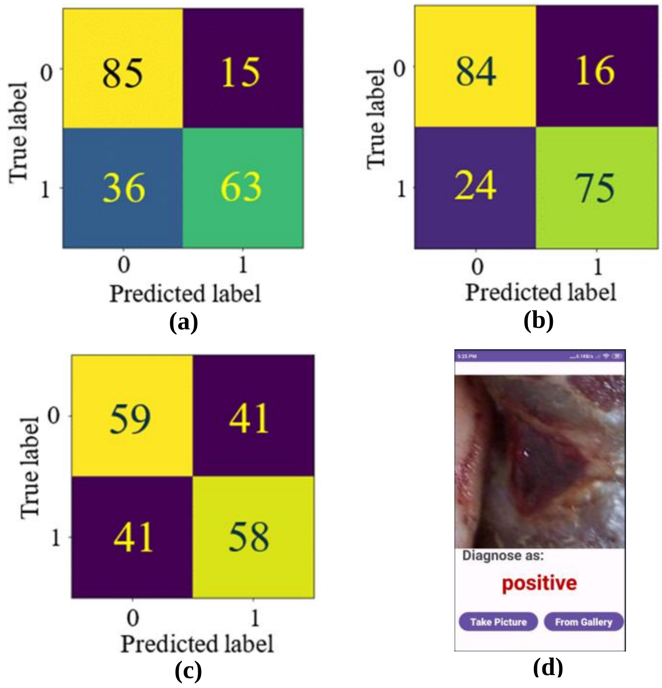


Fig. 6. Testing confusion matrix of (a) MobileNetV2, (b) EfficientNetB3, (c) Stack convolutional model, and (d) a screenshot of DFU image classification on Android smartphone. Class label 0: negative and 1: positive infection.

Table 1. Performance metrics of the target models optimized for DFU infection classification.

Model	**Performance metrics**				
	Accuracy	Sensitivity	Specificity	Precision	F1
MobileNetV2	0.744	0.636	0.85	0.81	0.72
EfficientNetB3	0.8	0.76	0.84	0.82	0.79
Stack model	0.588	0.586	0.59	0.586	0.586

4. Discussion

The key challenge in the early detection and diagnosis of DFU infection is the rigid scheduled care and screenings, which could be infrequent and unsystematic. This can lead to delayed treatment and result in further complications and increased morbidity. This work proposed using edge computing technology to establish more effective patient care and improve patients’ outcomes. This system promotes rapid classification of DFU infection, allowing an AI-assisted self-diagnosis using patient-captured images before they seek professional confirmation and prompt treatment from a health facility.

The developed mobile app tested on the testing images showed comparatively better classification performance achieved using the transfer learning methods on MobileNetV2 and EfficientNetB3 with a training accuracy of 0.74-0.8 than with the stack convolutional model (accuracy of 0.59). Unlike the stack model (consisting of 2.5 million learnable parameters that have never learned any features), the weights of the pretrained networks (i.e., 2.2 million in MobileNetV2 and 22 million in EfficientNetB3) were originally tuned on ImageNet data. The weights of these parameters have been efficiently updated and optimized for our two-class classification problem, enhancing the network's generalization capabilities.

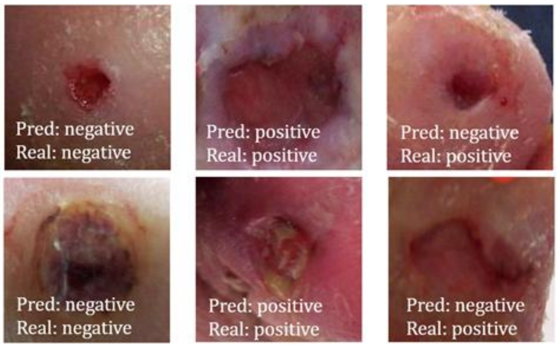
Both MobileNetV2 and the stack model are considerably lightweight, resulting in a faster convergence rate by two-fold than their deeper and wider EfficientNet counterpart. The efficiency of the latter architecture is also the reason for the best overall performance in Fig. 6. The results show that EfficientNetB3 can recognize positive infection cases with a lower misclassification rate of 24 % (and a superior sensitivity score of 0.76), as compared to 36 % in MobileNetV2, i.e., the sensitivity of 0.63. These models correctly detected most negative infection images, as can also be observed from the results in Fig. 7, producing a reasonable specificity of 0.85. The findings also show similar precision and F1-score performances, ranging between 0.72-0.82.

These scores in Table 1 may initially suggest that these pretrained networks overfitting to the negative infection training data, nonetheless the balanced dataset is used in this study, and this problem is not seen in the results of the stack model in Figs. 6(c) and 7. Therefore, we rule out the possibility of model bias. Images in both classes, as shown in Figs. 2 and 7, exhibit similar features in terms of colour and texture, except that the positive infection demonstrated greater impairment with increased exudate and wound depth, and sloughing and exposure of the dermal layer.

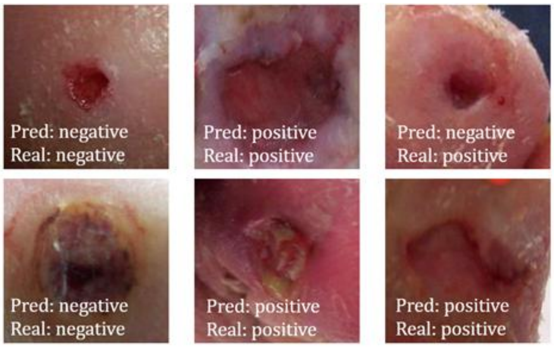
This complexity is further aggravated by significant variations in the appearance and characteristics of the wound bed, which vary in colour from yellow to black. Positive DFU infection images, whose wound bed is similar in hue to the surrounding skin, as shown on the rightmost bottom rows of Fig. 7, were misclassified by most models (i.e., MobileNetV2 and stack convolutional network) as noninfecting. Therefore, one possible reason for the misclassification rate is the infeasibility of the adopted networks in learning the high-level representations in the DFU infection dataset.

In addition, we do not exclude the possibility of poor image quality that degrades the model performance and the image's complex background that affects the analysis of colour and structure. Some images, e.g., seen in the centre image on the top row and the last image on the second row of Fig. 2, show insufficient image quality due to motion artifacts and the diversity of the image background. These cause uncertainties in the training process and possibly the learning of irrelevant information in the model training, which may contribute to the misclassification rate in the testing phase. Meanwhile, an example of background noise that considerably affects the models' performances is that shown on the rightmost image in the top rows of Fig. 7.

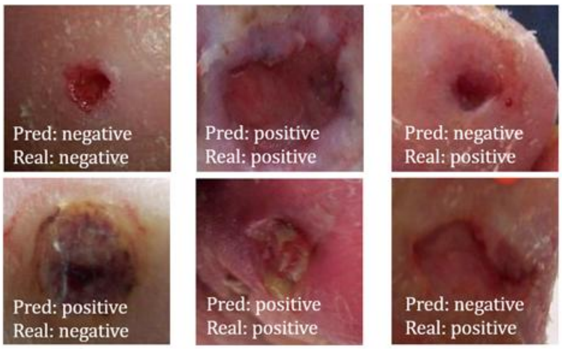
The dark background colour on the top right corner of this image (marked by the red box) could have been wrongly identified as the attributes of scabs, which are common in negative infection wounds (see leftmost image indicated by the white box in Fig. 7), causing misclassification of the image by the models. The proposed stack model performed inferiorly in all metrics evaluated, suggesting the need for a more robust weight updating strategy to enhance the accuracy. These experimental results show that the MobileNetV2 is superior in terms of computing and classification performances as compared to the competing models. This model can be easily adapted to the DFU image data and is effective for DFU infection detection.



(a) MobileNetV2



(b) EfficientNetB3



(c) Stack convolutional model

Fig. 7 Examples of prediction using the employed models tested on six images randomly chosen from the test set.

This work developed a Java-based user interface for serverless deployment and implementation of edge detection of DFU infection. This app can function with the device's CPU without needing an internet or a cloud server connection, providing a high degree of flexibility. Therefore, it is well suited for environments with inadequate ICT infrastructure. This process also results in faster response time and prediction speed. Since the testing stage of this app involved uploading images from the device gallery, the next step of this work is to validate the system's performance in real-world clinical settings using the image input from the device camera. Such an experiment is not possible in this study as the researchers had no access to the DFU patients.

In addition, there is room for improvement in the network's performance, which can be enhanced with deeper and wider network structures. More robust and effective image preprocessing steps may also be implemented to address the image quality issues due to artifacts, environment and working factors, such as poor ambient lighting conditions, imaging angle, and different camera settings and quality. Image segmentation could also be explored to remove the image background and integrate it into the system for more effective diagnosis.

5. Conclusion

This paper presents a complete working framework of an Android application for the classification of DFU infection status on the edge. This standalone system provides a solution to improving healthcare delivery by rapidly diagnosing a suspected wound and making the diagnosis process accessible to patients in remote care settings. This strategy allows patients to take prompt action in seeking further clinical opinions and treatment in cases of suspected positive infection, improving patient outcomes. This approach is also advantageous in time and cost-saving in terms of lower healthcare costs and resource utilization. The experimental results comparing the training and classification performances of MobileNetV2, EfficientNetB3, and a custom-developed stack convolutional model showed the superiority of MobileNetV2. The testing of this system showed an acceptable classification performance ranging from 0.63 to 0.85, which could be further improved by adopting more effective image preprocessing techniques and improving network architecture.

Acknowledgment

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Tier1 (Q381) and the Ministry of Higher Education (MOHE) through the Fundamental Research Grant Scheme (FRGS) (FRGS/1/2024/TK07/UTHM/02/5).

Abbreviations

ACC	Accuracy
CNN	Convolutional Neural Network
DFU	Diabetic Foot Ulcer
PREC	Precision
SENS	Sensitivity
SPEC	Specificity
TF	TensorFlow

References

1. Agarwal, A.; and Ojha, R. (2024). Prioritizing implications of industry-4.0 on the sustainable development goals: A perspective from the analytic hierarchy process in manufacturing operations. *Journal of Cleaner Production*, 444, 141189.
2. Tan, J.; Qin, F.; and Yuan, J. (2021). Current applications of artificial intelligence combined with urine detection in disease diagnosis and treatment. *Translational Andrology and Urology*, 10(4), 1769-1779.
3. Hussain, S. et al. (2022). Modern diagnostic imaging technique applications and risk factors in the medical field: A Review. *Biomedical Research International*, 2022(1), 1-19.
4. Stoumpos, A.I.; Kitsios, F.; and Talias, M.A. (2023). Digital transformation in healthcare: Technology acceptance and its applications. *International Journal of Environmental Research and Public Health*, 20(4), 1-44.
5. Gaube, S.; Biebl, I.; Engelmann, M.K.M.; Kleine, A.; and Lermer, E. (2024). Comparing preferences for skin cancer screening: AI-enabled app vs. dermatologist. *Social Science and Medicine*, 349, 116871.
6. Singh, V. (2022). *Digital and personalized healthcare system for COVID-19 and future pandemics*. In *Pandemic detection and analysis through smart computing technologies*. Apple Academic Press.
7. Ahmed, S.; Yong, J.; and Shrestha, A. (2023). The integral role of intelligent IoT system, cloud computing, artificial intelligence, and 5G in the user-level self-monitoring of COVID-19. *Electronics*, 12(8), 1912.
8. Hemdan, E.E.D.; El-Shafai, W.; and Sayed, A. (2023). CR19: A framework for preliminary detection of COVID-19 in cough audio signals using machine learning algorithms for automated medical diagnosis applications. *Journal of Ambient Intelligence and Humanized Computing*, 14(9), 11715-11727.
9. Arntz, A. et al. (2023). Technologies in home-based digital rehabilitation: Scoping review. *JMIR Rehabilitation and Assistive Technologies*, 10, e43615.
10. Lyttle, B.D. et al. (2023). Effects of microRNAs on angiogenesis in diabetic wounds. *Frontiers in Medicine*, 10, 1-15.
11. Yap, M.H. et al. (2021). Development of diabetic foot ulcer datasets: An overview. *Diabetic Foot Ulcers Grand Challenge*, 13183, 1-18.
12. Raja, J.M.; Maturana, M.A.; Kayali, S.; Khouzam, A.; and Efeovbokhan, N. (2023). Diabetic foot ulcer: A comprehensive review of pathophysiology and management modalities. *World Journal of Clinical Cases*, 11(8), 1684-1693.
13. Huerta, C.T.; Voza, F.A.; Ortiz. Y.Y.; Liu, Z.J.; and Velazquez, O.C. (2023). Mesenchymal stem cell-based therapy for non-healing wounds due to chronic limb-threatening ischemia: A review of preclinical and clinical studies. *Frontiers in Cardiovascular Medicine*, 10, 1-13.
14. Jneid, J.; Lavigne, J.P.; La Scola, B.; and Cassir, N. (2017). The diabetic foot microbiota: A review. *Human Microbiome Journal*, 5, 1-6.
15. Joachim, S.; Forkan, A.R.M.; Jayaraman, P.P.; Morshed, A.; and Wickramasinghe, N. (2022). A nudge-inspired AI-Driven health platform for self-management of diabetes. *Sensors*, 22(12), 4620.
16. Tripathi, A.K.; Mishra, S.; and Vasudevan, S.K. (2023). Smart diabetic prediction: An intelligent IoT-based diabetic monitoring system with stacked

spatio temporal features-based multiscale dilated deep temporal convolutional network. *Sensing and Imaging*, 25(1), p.2.

17. Shen, J. et al. (2020). An innovative artificial intelligence-based app for the diagnosis of gestational diabetes mellitus (GDM-AI): Development study. *Journal of Medical Internet Research*, 22(9), e21573.
18. Indraswari, R.; Rokhana, R.; and Herulambang, W. (2022). Melanoma image classification based on MobileNetV2 network. *Procedia Computer Science*, 197, 198-207.
19. Rashid, J. et al. (2022). Skin cancer disease detection using transfer learning technique. *Applied Science*, 12(11), 1-16.