Diabetic Foot Ulcer Ischemia and Infection Classification Using EfficientNet Deep Learning Models

Ziyang Liu ⓘ, Josvin John ⓘ, and Emmanuel Agu ⓘ

Abstract—Motivation: Infection (bacteria in the wound) and ischemia (insufficient blood supply) in Diabetic Foot Ulcers (DFUs) increase the risk of limb amputation. *Goal:* To develop an image-based DFU infection and ischemia detection system that uses deep learning. *Methods:* The DFU dataset was augmented using geometric and color image operations, after which binary infection and ischemia classification was done using the EfficientNet deep learning model and a comprehensive set of baselines. *Results:* The EfficientNets model achieved 99% accuracy in ischemia classification and 98% in infection classification, outperforming ResNet and Inception (87% accuracy) and Ensemble CNN, the prior state of the art (Classification accuracy of 90% for ischemia 73% for infection). EfficientNets also classified test images in a fraction (10% to 50%) of the time taken by baseline models. *Conclusions:* This work demonstrates that EfficientNets is a viable deep learning model for infection and ischemia classification.

Index Terms—Deep Learning, Diabetic Foot Ulcers, EfficientNet, Infection, Ischemia.

Impact Statement— The EfficientNet neural networks architecture was innovatively adapted for ischemia and Infection recognition in Diabetic Foot Ulcer Images and significantly outperformed the state of the art in both accuracy and speed.

I. INTRODUCTION

OVER 6.5 million people in the US (or approximately 2% of the population [1]) have chronic wounds, which are prevalent in the elderly population [2], [3] and cost the healthcare system over $25 billion annually [4]. Chronic wounds are often painful and require proper treatment including regular cleaning, debridement, changing of dressings and the use of antibiotics in order to heal properly [5]. Due to the large and growing number of chronic wounds, there is an increasing demand for information technology solutions that assist the work of medical personnel, improve efficiency and reduce the cost of care.

There are four main types of chronic wounds: pressure ulcers, diabetic foot ulcers, venous ulcers, and arterial ulcers. Diabetic foot ulcers (DFUs), the focus of this paper, are a major complication of diabetes. Infection and ischemia are two common problems that occur in the healing process of DFUs, which may lead to amputation of limbs and admission of patients into hospitals [6]. After limb amputation, the patient's quality of life degrades quickly with a life expectancy of fewer than 3 years thereafter [7]. Infection occurs in 40%–80% [8] of DFUs and ischemia occurs in almost 50% of DFUs [9]. Infection is caused by the presence of bacteria in the wound, which causes cell death. DFUs are at risk of becoming infected as they are often located on the lower limbs such as on the sole of the patient's foot [10]. Ischemia is caused by insufficient blood circulation due to chronic complications of diabetes. It is essential to recognize infection and ischemia of DFUs early to reduce the risk of amputation. The economic burden of treating DFUs including infection and ischemia, ranges from $9–13 billion [11].

To diagnose infection and ischemia accurately, the wound bacteriology needs to be tested combined with review of patient records such as clinical history, physical health and blood tests. But this information is not always available to clinicians. Experienced wound experts are able to recognize the presence of infection and ischemia in DFUs by visual inspection. Visual cues of wound infection include increased redness in and around the ulcer, and colored purulent. Visual cues for a wound with ischemia include the presence of poor reperfusion to the foot, or black gangrenous toes (tissue death to part of the foot). Prior work has demonstrated that infection and ischemia can be detected reliably from the appearance of the wound in a photograph using image analyses methods, without requiring wound tests and accurate records [12]. The goal of this work is to create deep learning models that can replicate the diagnoses of infection and ischemia in DFUs by wound experts. Such models can be used remotely and especially in situations (e.g. patients' homes) where wound experts are not available. The accuracy of medical image recognition tasks using machine learning algorithms has increased significantly in recent years, outperforming and producing more consistent assessments than skilled humans
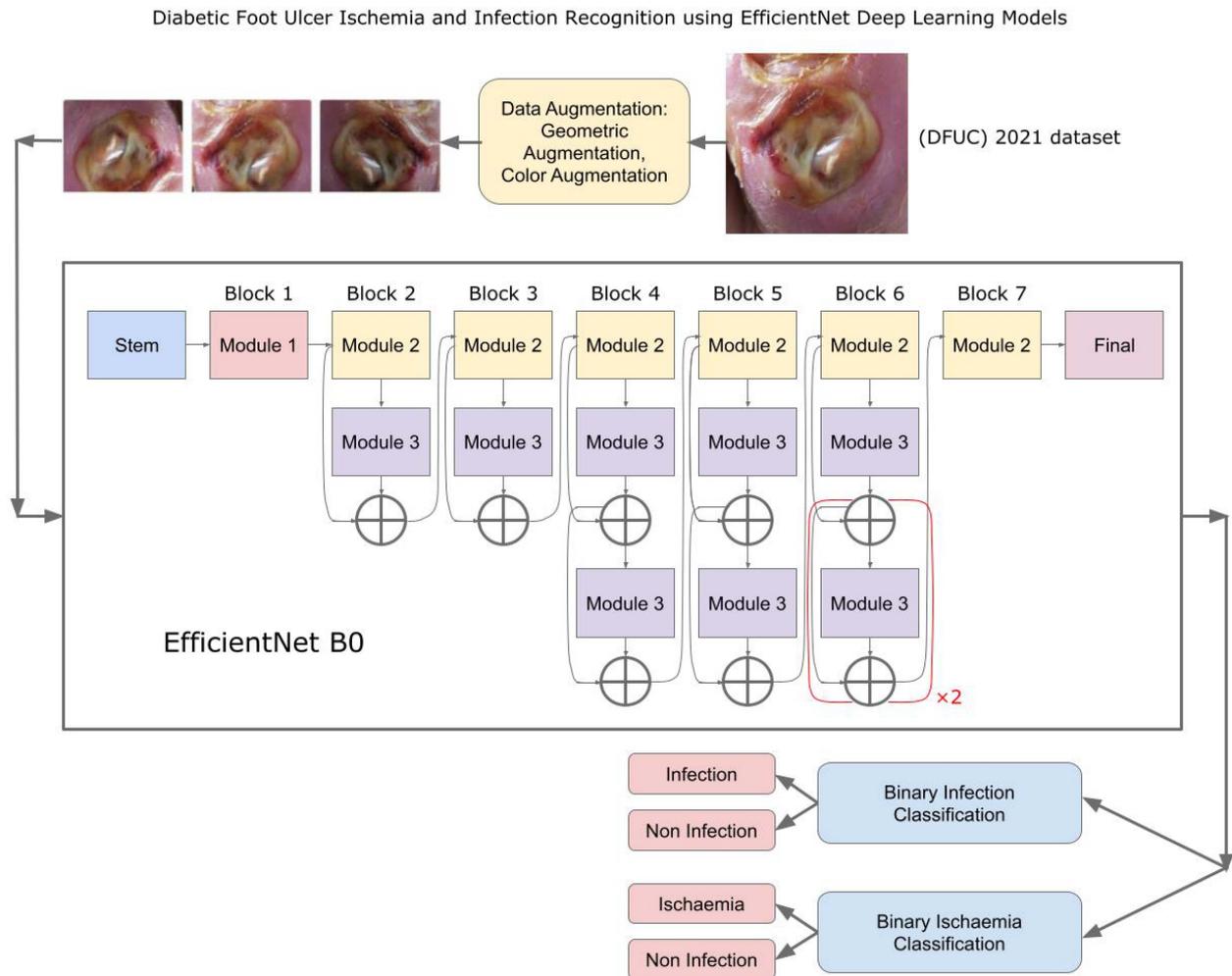
Fig. 1. Diabetic Foot Ulcer Ischemia and Infection Recognition using EfficientNet Deep Learning Models.

including clinicians in some cases [13], [14]. Moreover, specific to wounds, Netten et al. [15] showed that clinicians achieved low validity and reliability for remote assessment of DFU mobile phone images.

Our approach: In this work, a DFU infection and Ischaemia detection system that utilizes the EfficientNets neural networks architecture (See Fig. 1) for binary classification of wound images, was developed. The research utilized a wound infection and ischemia dataset made publicly available as part of the Diabetic Foot Ulcers Grand Challenge (DFUC) 2021 [16], [17] with details expounded on in Section II-A. EfficientNets [18], [19] are computationally efficient with fewer parameters than other methods and have achieved 84.4% top-1 state of the art results on the ImageNet dataset for classifying 14 million natural images into 1,000 categories. To achieve computational efficiency, a key innovation in EfficientNets is compound scaling, a new scaling method in which a compound coefficient is used to uniformly scale the depth, width and resolution of a neural network. Figure 3(a) shows various methods of scaling including compound scaling. The intuition behind compound scaling is that the network needs more layers to increase the receptive field and more channels to capture more fine-grained

patterns in larger images. Wound image analysis is a fine-grained image analysis problem [20], wherein images belonging to distinct target categories appear quite similar. EfficientNet achieves state-of-the-art performance by scaling up the CNN-based model in order to detect wound details in higher resolution images.

Challenges: In DFU assessment using deep learning methods include: (1) High inter-class similarity and intra-class variations of the infection and ischemia wound classes in DFU images; (2) Highly variable and non-standardized DFU dataset imaging conditions with large variations in the camera's distance from the foot, its orientation (pose) and lighting conditions; (3) Lack of patient demographic information such as patient ethnicity, age, sex, foot size or any accompanying meta-data for the DFU dataset.

Prior work: Earlier approaches to recognize infection in wounds utilized traditional machine learning techniques with handcrafted features [21], [22]. Hsu et al. [21] utilized a clustering technique followed by machine learning infection classification using the Support Vector Machines (SVM) algorithm. Hsu et al. [22] first segmented the wound image and then recognized wound infection using the SVM algorithm.
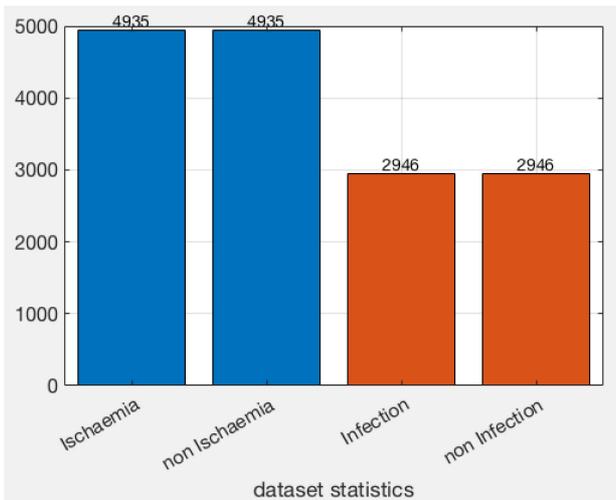
(a) Statistics of the DFUC 2021 dataset utilized in this work

(b) 3 augmentations of a wound image. Left: affine transformations such as shearing and scaling, Middle: increase in brightness, Right: a center cropped image

(c) 3 augmentations of a wound image. Left: image flipped horizontally, Middle: image flipped vertically, Right: the hue changed by 50%

(d) 3 augmentations of a wound image. Left: image rotated clockwise, Middle: the image rotated by an angle, Right: the saturation increased by 50%

Fig. 2. Dataset Statistics and Examples of data augmentation of a wound image.

As manual feature engineering is tedious, time-consuming and error-prone, more recent work has utilized neural networks approaches [23], [24], [25] that generally have superior performance when adequate data is available. Wang et al. [23] used a novel variant of a deep ConvNet for wound segmentation and then detected infection using SVM. Nejati et al. [24] used a pre-trained AlexNet neural network as a feature extractor, then performed feature reduction using Principal Components Analysis (PCA) and classification using linear SVM on various wound tissues including infected tissue. Goyal et al. [25] proposed a novel DFUNet architecture which combines depth and parallel convolution layers to discriminate the DFU from

healthy skin. Goyal et al. [12] achieved the highest accuracy of 90% accuracy for binary ischemia classification and 73% for binary infection classification using the Ensemble CNN on the same DFUC2021 dataset analyzed in this paper [16], [17]. Their neural networks model works by using the Inception-V3, ResNet50, and InceptionRes-NetV2 CNN models to extract bottleneck features from wound images, which were then combined and classified using an SVM classifier [12]. Yap et al. [16] utilized the VGG16, ResNet101, InceptionV3, DenseNet121 and EfficientNet architectures to classify the DFUC2021 dataset as a four-class classification (infection (yes/No) and ischaemia (yes/no)) problem [16]. The EfficientNet B0 achieved the best results for four-class classification with macro-average Precision, Recall and F1-Score of 0.57, 0.62 and 0.55 respectively. Al-Garaawi et al. [26] used a CNN-based DFU classification method to classify Infection and Ischaemia in the DFUC dataset as a 2-class classification problem [16], [17]. The proposed 2-stage method consisted of first using mapped binary patterns to extract texture information before feeding the stack of RGB and mapped binary pattern images to the CNN. This method achieved 0.995% (AUC) and 0.990% (F-measure) for Ischaemia, and 0.820% (AUC) and 0.744% (F-measure) for infection. It is important to note that the three works [12], [16], [26], which explored multi-class classification on the DFUC2021 dataset were by the group that released the dataset. However, the publicly released version of the DFUC2021 dataset [16], [17] that we analyzed, only included labels that support binary classification. Based on this limitation, we were only able to explore binary classification in this paper but have the goal of labeling and exploring multiple classes in future. Table 1 summarizes prior machine learning based wound infection and and ischaemia work.

Our proposed approach is related to, but outperforms the work of Goyal et al. [12]. Rather than using a single CNN-based model for four-class classification of Infection (Yes/No) and Ischaemia (Yes/No) as was done by Yap et al. [16], in order to achieve the best possible performance, we trained two separate CNN-based models to classify Infection and Ischaemia respectively. Also, while prior work [12], [26] benchmarked the accuracy of various proposed CNN-based models this is the first work to benchmark and establish the superior classification speed of the EfficientNet model that will facilitate faster diagnoses in practical deployments.

Our contributions: in this paper include:
1) Envisioning and proposing an innovative DFU Infection and Ischaemia detection system that consists of two EfficientNets, one each for Infection and Ischaemia. This envisioned system can be utilized for remote wound assessment in the patient's home. A visiting nurse would upload a wound picture taken in the patient's home to a cloud server for assessment of infections and ischemia. The proposed EfficientNet models will analyze the wound image in the cloud. The nurse can use the proposed wound system's predictions as supporting evidence for treating the patient.
2) Innovatively adapting the EfficientNet neural networks architecture for analyses of fine-grained wound details
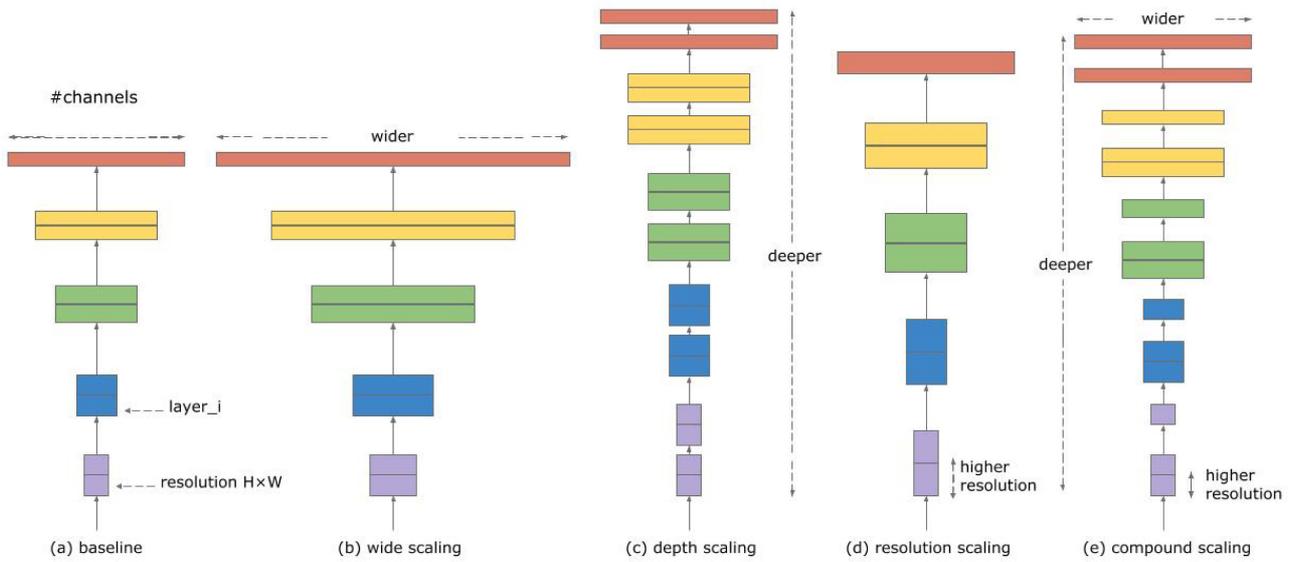
(a) Various Neural Networks Model Scaling Methods. a: baseline network; b to d: conventional scaling that only increases one of width, depth, or resolution in the network; e: compound scaling method that uniformly scales all three dimensions with a fixed ratio
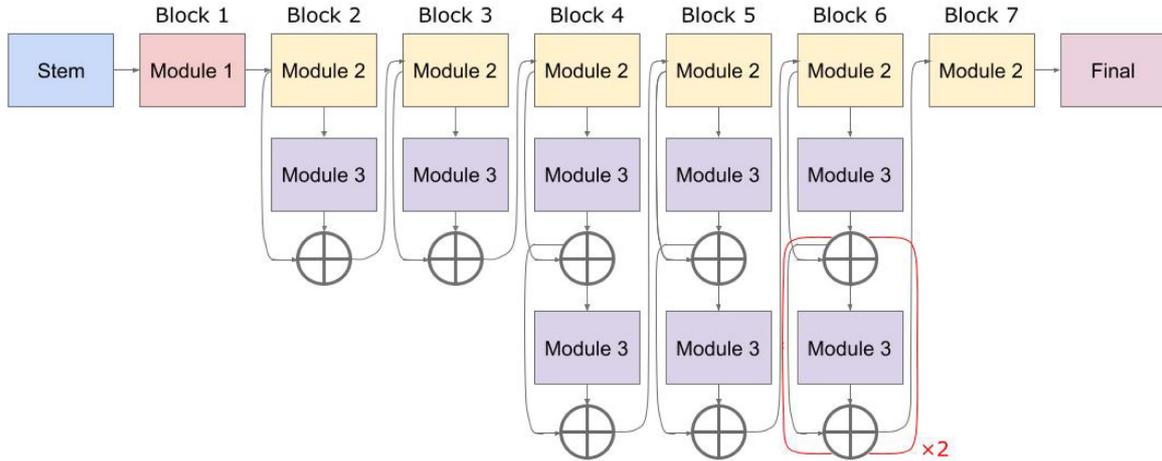
(b) EfficientNet B0 Architecture
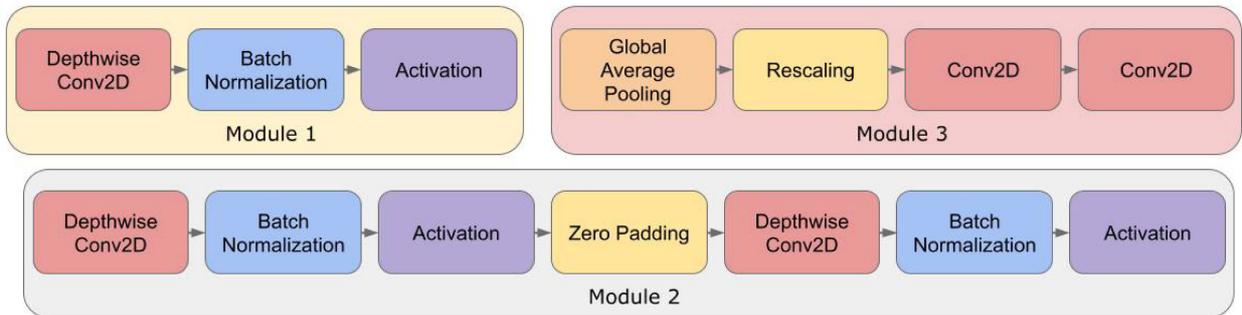
(c) Modules in EfficientNet (Module 1 and Module 2 are two types of Inverted Residual Blocks shown in Figure 4b)

Fig. 3.   EfficientNet scaling and EfficientNet Architecture.

TABLE I
PRIOR WOUND INFECTION AND ISCHAEMIA RECOGNITION WORK AND DFUC2021 DATASET STATISTICS

(a) Prior work on DFU Classification and Wound Infection Recognition

Authors/ Citation	Specific Machine Learning problem	Summary of Approach	No. of target classes	Best performance results
Wound Ischaemia and Infection Recognition using deep neural network models				
Goyal et al. 2020 [12]	DFU wound Ischaemia and Infection Recognition	Ensemble CNN	2 classes (Ischaemia: Yes/No; Infection: Yes/No)	accuracy (Ischaemia: 90%, Infection: 73%)
Yap et al. 2021 [16]	DFU wound Ischaemia and Infection Recognition	VGG16, ResNet101, InceptionV3, DenseNet121, EfficientNet	4 classes (both Infection and Ischaemia, Infection, Ischaemia, None)	EfficientNet B0: macro-average Precision, Recall and F1-Score of 0.57, 0.62 and 0.55
Al-Garaawi et al. 2022 [26]	DFU wound Ischaemia and Infection Recognition	CNN-based DFU classification method	Part A: 2 classes (healthy skin and DFU); Part B: 2 classes (Ischaemia: Yes/No; Infection: Yes/No)	Ischaemia: 0.995% (AUC), 0.990% (F-measure) Infection: 0.820% (AUC), 0.744% (F-measure)
Wound Infection Recognition using deep neural network models				
Wang et al. 2015 [23]	wound segmentation and infection detection	deep neural network, SVM	2 classes (infection and no infection)	infection classification accuracy 95.6%
Nejati et al. 2018 [24]	classification of 7 tissue types including infection	deep neural network, SVM	7 classes (Necrotic, Healthy Gran, Slough, Infected, Unhealthy Gran, Hyper Gran, Epithelialization)	tissue classification accuracy 86.4%
Wound Infection Recognition using traditional machine learning techniques				
Hsu et al. 2017 [21]	detection of 4 tissue types including infection	clustering method and classification using SVM	4 classes (Swelling, Blood Region, Infected, Tissue Necrosis)	detection accuracy of 95.23%
Hsu et al. 2019 [22]	wound segmentation and infection detection	robust image segmentation, classification using SVM	4 classes (Swelling, Granulation, Infection, Tissue Necrosis)	tissue classification accuracy 83.58%

(b) Statistics of different versions of the DFUC2021 dataset

Authors/ Citation	Specific Machine Learning problem	No. of target classes	Statistics of dataset
Goyal et al. 2020 [12]	DFUC2021 dataset classification	2 classes (Ischaemia: Yes, No; Infection: Yes, No)	Ischaemia: (Yes, 235; No, 1431) augmented to (Yes, 4935; No, 4935); Infection: (Yes, 982; No, 684) augmented to (Yes, 2946; No, 2946)
Yap et al. 2021 [16]	DFUC2021 dataset classification	4 classes (both Infection and Ischaemia, Infection, Ischaemia, None)	both Infection and Ischaemia: 621; Infection: 2555; Ischaemia: 277; none of them: 2552
Al-Garaawi et al. 2022 [26]	DFU dataset classification and DFUC2021 dataset classification	Part A: 2 classes (healthy skin and DFU); Part B: 2 classes (Ischaemia: Yes, No; Infection: Yes, No)	Part A: 641 healthy, 1038 Ulcer Part B: Ischaemia: augmented (Yes, 4935; No, 4935); Infection: augmented (Yes, 2946; No, 2946)
Goyal et al. 2018 [25]	DFU dataset classification	2 classes (healthy skin and DFU)	641 healthy, 1038 DFU

in high-resolution images for binary classification of: (1) ischemia vs. No ischemia; and (2) infection vs. no infection.

3) Performing rigorous evaluations of the EfficientNet architecture, demonstrating that it achieves 99% accuracy in ischemia classification and 98% in infection classification, outperforming various baseline image classification CNNs models including DenseNet121, ResNet50 and Inception V3. It also outperformed Ensemble CNN [12], the previous state-of-the-art (Classification accuracy of 90% for ischemia and 73% for infection) with a slightly higher accuracy in the classification of ischemia than for infection recognition.

4) Evaluation of wound image analysis speed. Due to compound scaling, the EfficientNets model was significantly faster, running in 10% to 50% (About 5 minutes) of the time taken by baseline models (About 1 h). The EfficientNet running time is reasonable for the envisioned usage scenario.

II. MATERIALS AND METHODS

A. DFU Infection and Ischemia Dataset

The Diabetic Foot Ulcers Grand Challenge (DFUC) 2021 dataset [16], [17] contained images of DFUs that were collected from the Lancashire Teaching Hospital with the approval for research from the U.K. National Health Service (NHS) Research Ethics Committee (REC) (NHS REC reference no.

15/NW/0539). Statistics for the DFUC are showing Fig. 2(a). Each class had about 3000 images captured in stable room lighting at a distance of 30-40 cm from the foot with the plane of the ulcer parallel to the image. The images were acquired by a podiatrist and a diabetic ulcers consultant physician, both with more than 5 years of professional experience, who produced ground truth labels on infection and ischemia status. The original DFU images had sizes varying between 1600×1200 and 3648×2736, which were resized to a dimension of 640×640 that was suitable for deep learning, optimizes performance and minimizes computational costs.

B. Methodology

Gathering and annotating the large number of medical images for deep learning tasks can be tedious and expensive. To reduce overfitting and improve model robustness, transfer learning and data augmentation of the original DFU dataset were utilized. For transfer learning, the EfficientNet deep learning model was first pre-trained on the large ImageNet dataset. The weights the model learned from Imagenet were then utilized as initial weights to train it on the smaller DFU dataset. For data augmentation, slightly modified versions of each original wound image were generated and added to the dataset, increasing its size while preserving important semantic information. The proposed approach, which is now described in some detail, includes the data augmentation techniques and the EfficientNet deep learning architecture used for ischemia and infection classification.

(a) Data Augmentation scenarios

ID	Name	Description
A	Saturation, Contrast, Brightness and Hue	Modify saturation, contrast, and brightness by random factors sampled from a uniform distribution of [0.7; 1.3], Also, shift the hue by a value sampled from a uniform distribution of [-0.1; 0.1], simulating changes in color due to camera settings and lesion characteristics.
B	Affine	Rotate the image by up to 90, shear by up to 20, and scale the area by [0.8; 1.2]. New pixels are filled symmetrically at edges. This can reproduce camera distortions and create new lesion shapes.
C	Flips	Randomly flip the images horizontally and/or vertically
D	Random Crops	Randomly crop the original image. The crop has [0.4 - 1.0] of the original area, and 3/4 - 4/3 of the original aspect ratio.
E	Combination Set	Combination of D → B → C → A

(b) EfficientNet-B0 baseline network – Each row shows a stage i with $\hat{L}_i$ layers, input resolution $(\hat{H}_i, \hat{W}_i)$ and output channels $\hat{C}_i$

Stage	Operator	Resolution	Channels	Layers
i	$\hat{F}_i$	$\hat{H}_i \times \hat{W}_i$	$\hat{C}_i$	$\hat{L}_i$
1	Conv3×3	224 × 224	32	1
2	MBConv1, k3×3	112 × 112	16	1
3	MBConv6, k3×3	112 × 112	24	2
4	MBConv6, k5×5	56 × 56	40	2
5	MBConv6, k3×3	28 × 28	80	3
6	MBConv6, k5×5	14 × 14	112	3
7	MBConv6, k5×5	14 × 14	192	4
8	MBConv6, k3×3	7 × 7	320	1
9	Conv1×1 & Pooling & FC	7 × 7	1280	1

1) Data Augmentation Techniques: Data augmentation operations utilized were grouped into the five types shown in Table 2(a), selected because they had produced good results for the similar skin lesion analysis task [27], [28], [29]. The operations included color modifications and geometric transforms (rotation, scaling, random cropping), image processing operations and combinations. Figure 2(b), (c), and (d) show examples of modified (augmented) wound images from the infection dataset.

2) Deep Learning Architectures for Binary Ischemia and Infection Classification: included pre-trained, fine-tuned, CNN models (transfer learning) including VGG16 [30], Inception-v3 [31], ResNet50 [32], DenseNet121 [33], EfficientNet B0-B7 [18], [19]. Additional details on these models will be provided in latter sections. To perform transfer learning, the weights of the first few layers of the pre-trained networks that typically learn common features, such as textures, edges and curves in the image were frozen. The latter layers of the neural network, which focus on learning task-specific features, were re-trained.

3) EfficientNet Model Architecture and Compound Scaling: Compound scaling is one of the main innovations in the EfficientNet architecture. The scaling problem can be formulated as follows [18]. A ConvNet Layer i can be defined as a function: $Y_i = F_i(X_i)$, where F_i is the operator and Y_i is the output tensor. X_i is the input tensor with shape $< H_i, W_i, C_i >$, where H_i and W_i are spatial dimensions and C_i is the channel dimension. The ConvNet is defined as:

$$N = \bigodot_{i=1...s} F_i^{L_i}(X_{<H_i, W_i, C_i>}) \qquad (1)$$

Here $F_i^{L_i}$ means that layer F_i is repeated L_i times in stage i and $< Hi, Wi, Ci >$ represents the shape of input tensor X of layer i. Regular ConvNet designs focus on finding the best layer architecture F_i but model scaling tries to expand the network length L_i, width C_i, and/or resolution (H_i, W_i) with F_i remaining the same in the pre-defined baseline network. Conventional model scaling methods scale ConvNets in one of the d, w, r dimensions as their optimal values depend on each other. The model's goal is to maximize accuracy subject to any resource constraints. This problem can be formulated as the optimization problem in equation 2, where w, d, r are coefficients for scaling network width, depth, and resolution and F_i, L_i, H_i, W_i, C_i are pre-defined parameters in the baseline network.

$$\max_{d,w,r} \quad Accuracy(N(d,w,r))$$
$$\text{s.t.} \quad N(d,w,r) = \bigodot_{i=1...s} \hat{F}_i^{d \cdot \hat{L}_i}(X_{<r \cdot \hat{H}_i, r \cdot \hat{W}_i, w \cdot \hat{C}_i>})$$
$$Memory(N) \leq target\_memory$$
$$FLOPS(N) \leq target\_flops \qquad (2)$$
$$depth : d = \alpha^\phi$$
$$width : w = \beta^\phi$$
$$resolution : r = \gamma^\phi$$
$$\text{s.t.} \quad \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$$
$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1 \qquad (3)$$

The compound scaling method applied to EfficientNet involves uniformly scales its width, depth, and resolution using the same compound coefficient ϕ as expressed in Equation (3). Optimal values of the α, β, γ parameters can be determined using grid search. Variants B1 to B7 of the EfficientNet model shown in Table 2, are created by scaling the baseline model EfficientNet-B0 using various compound coefficients ϕ in Equation (3). For EfficientNet-B0, ϕ is set to 1 and optimal parameter values are $\alpha = 1.2$, $\beta = 1.1$ and $\gamma = 1.15$ subject to the constraints in Equation 3. In the next step, α, β, γ are set as constants.

The main building block of the EfficientNet architecture is the mobile inverted bottleneck block [34], [35] or inverted residual block, which is denoted as MBConv in stages 2 to 8 in Table 2. These blocks also apply squeeze and excitation [36] optimizations as well as the swish activation [37], which is shown in Figure 4(c) and Figure 4(d). Detailed technical details of the inverted residual block and the Squeeze-and-excitation (SE) block and Swish Activation are in the Supplementary Materials.

Figure 3(b) illustrates EfficientNet B0. Its blocks correspond to the stages of the EfficientNet-B0 architecture in Table 2 and

(a) Residual Block

(b) Inverted Residual Block

(c) SE-ResNet Module
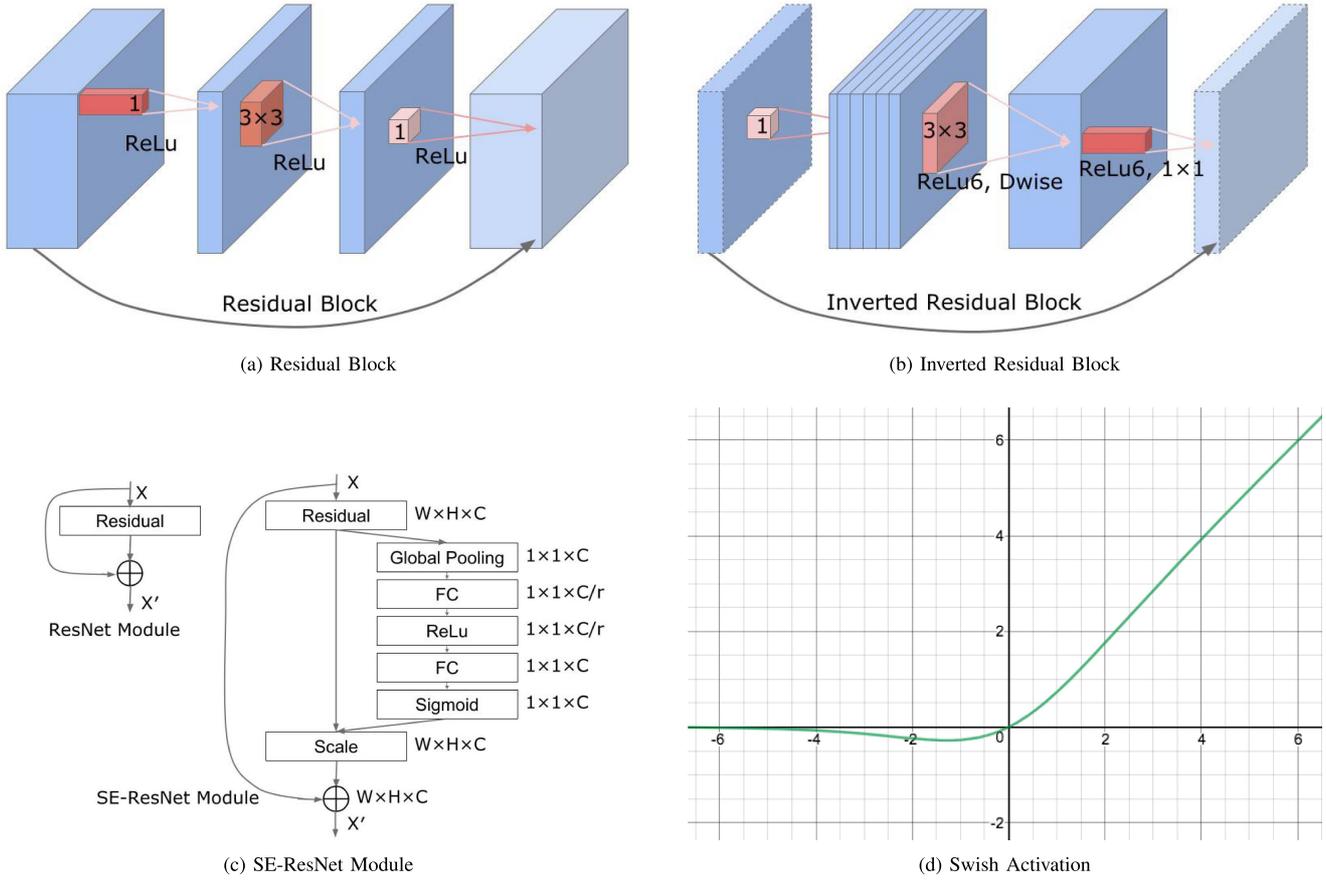
(d) Swish Activation

Fig. 4. Residual Block, Inverted Residual Block, SE-ResNet Module and Swish Activation.

Figure 3(c) provides more details on its modules shown in Figure 3(b). As EfficientNets B1 to B7 have a similar architecture, additional details will be suppressed due to space constraints. Figure 4(b) illustrates the inverted residual block, which will be explained in more detail in the Supplementary Materials. EfficientNet architectures decrease the number of trainable parameters significantly, yielding training and testing times as well as memory usage that are much smaller than previous models. The squeeze and excitation blocks in EfficientNets scale each channel of the model based on its importance, enabling it to focus on useful information. Consequently, the EfficientNets architecture improves on the accuracy, training and test times of similar models for fine-grained classification tasks such as wound infection and ischemia recognition.

C. Evaluation Metrics

EfficientNet as well as a comprehensive set of baseline models were evaluated using the following performance metrics:

True Positive Rate (TPR), also known as recall or Sensitivity:

$$TPR = \frac{TP}{TP + FN}$$

Precision:

$$Precision = \frac{TP}{TP + FP}$$

False Positive Rate (FPR):

$$FPR = \frac{FP}{FP + TN}$$

Specificity:

$$FPR = \frac{TN}{TN + FP}$$

Accuracy:

$$ACC = \frac{TP + TN}{P + N}$$

Other metrics include F1-Score, which is the Harmonic Mean of Precision and Recall, the Receiver Operating Characteristic (ROC), which is the plot of TPR (sensitivity) versus FPR (1-specificity), and the AUC, which stands for "Area under the ROC Curve." That is, AUC measures the entire two-dimensional area below the entire ROC curve from (0,0) to (1,1). An ideal classifier should have 100% sensitivity (TPR = 1) and 100% specificity ($TNR = 1$, or $FPR = 0$)

D. Baseline Neural Networks Models

1) VGG16: is a 16-layer CNN architecture proposed by Simonyan et al. [30], which achieved the best performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

2014, classifying 1 million images into 1000 possible classes with a top-5 error rate of 7.3%.

2) ResNet50: or deep residual network proposed by He et al.,he2016deep and introduced the Residual Block that alleviates the vanishing gradient problem when training very deep neural networks. The Residual Block has skip connections between layers in the ResNet architecture, enabling deep neural networks that have hundreds or even thousands of layers to achieve compelling performance.

3) Inception v3: is a neural networks architecture proposed by Szegedy et al. [31], [38], which introduced multiple innovations to modify previous versions of the Inception architecture in order to reduce computational cost. Specific techniques proposed to reduce the computational cost associated with scaling of inception v3 included factorized convolutions, regularization, dimension reduction, and parallelized computations.

4) DenseNet121: or Dense Connected Convolutional Network (DenseNet) was proposed by Huang et al. [33]. DenseNet has feed-forward connections from each layer to every other layer that was based on the observation that CNNs become more accurate and more efficient to train when connections between layers are shorter. The structure of DenseNet can mitigate the vanishing-gradient problem, increase feature propagation and feature reuse, as well as decreasing the number of parameters in the model [33].

5) Ensemble CNN: is the state-of-the-art neural networks model for classifying wound infection and ischemia [12]. It uses three neural network architectures (Inception-V3, ResNet50, and InceptionRes-NetV2) to extract bottleneck features from DFU images that were combined before being fed into an SVM classifier that performs ischemia and infection classification.

E. Experimental Methodology

The infection and ischemia datasets were first split into 70% training, 15% validation and 15% testing sets. Data-augmentation operations were then applied to each original wound image to augment the training, validation and test sets for infection and ischemia classification. By augmenting the 5890 original foot infection images in the infection dataset, 40740 images (training), 8730 images (validation), and 8730 images (testing) were generated. From the 9870 original foot ischemia images, after augmentation, there were 6909 images (training), 1480 images (validation), and 1481 images (testing). Before classification by various CNN models, each wound image was resized to a dimension of 120×120 RGB images for the VGG16 [30] model and a dimension of 224×224 RGB images for the ResNet50 [32], DenseNet121 [33], and EfficientNet B0-B7 [18] architectures. For the Inception-v3 [31] model, the images were resized to a dimension of 299×299. PyTorch 1.8.1 was used as the framework for performing these experiments.

III. RESULTS

Overall performance: Tables 3(a) and 3(b) show results of the evaluation of EfficientNet and various baseline models using various metrics including Accuracy, F1-Score, AUCROC (Area under the ROC Curve), Sensitivity, Specificity, True Positive

Rate (TPR), False Positive Rate (FPR). The models generally performed better in the binary classification of ischemia than infection. The average accuracy of all the models in the ischemia dataset was 97.636%, which was notably better than the average accuracy of 92.89% in the infection dataset. EfficientNet classified both infection and ischemia better than baseline models. The EfficientNet model B5 achieved a Sensitivity of 99% for classifying infection, which was 4% higher than 95% achieved by Dense-Net121, the second-best model. For classifying ischemia, the EfficientNet model B1 achieved a Sensitivity of 100%, which was similar for the Inception V3 model and 4% higher than 96% achieved by DenseNet121. The EfficientNet model B1 also achieved the highest Specificity of 99%, which was 1% higher than that achieved by DenseNet121 (98%). EfficientNet B1 also had an AUC of 0.9939, outperforming all other baselines including the Ensemble CNN the previous state-of-the-art for wound infection and ischemia classification [12].

Performance of baseline models: DenseNet121 was generally the best performing baseline with Sensitivity and Specificity of 95% and 94% respectively for infection classification. For ischemia classification, Inception V3 had the highest Sensitivity metric of 100%, whereas DenseNet121 had the highest Specificity of 98%.

Test time: In the envisioned use case, a visiting nurse takes a picture of the patient's wound in their home, submits it to our infection and ischemia classification system, and has to wait for its results before taking action. This makes it important that the model has a reasonable test time. As shown in Tables 3(a) and 3(b), due to compound scaling, the EfficientNet models had faster test times than all baselines. For infection classification, the EfficientNet models' testing times were approximately 1/2 to 1/10 of that of baseline models. For ischemia classification, the EfficientNet models' testing times were less than 1/10 of that of baseline models. In order to ensure a fair comparison, all models were trained and tested on the same desktop machine with an NVIDIA GTX 1080 Ti GPU.

Performance gains due to data augmentation: were evaluated by training the EfficientNet B5 model with and without data augmentation for infection classification with results shown at the bottom of Table 3(a). The data augmentation in scenario (E) performed best. It combines geometric and color transformations, surpassing the AUC values achieved by the other augmentation scenarios, no augmentation or individual augmentations. Data augmentation significantly improved the performance of EfficientNet B5 for infection classification but did not improve the performance of EfficientNet B1 on the ischemia dataset, possibly because its performance on the ischemia dataset was already high.

Exploring mis-classifications: Figure 5(c), (d), (g), and (h) show examples of images from the infection and ischemia datasets, which were mis-classified by the EfficientNet model. Additional examples of mis-classification by the EfficientNet model are also shown in Figure 6(c). These mis-classifications occurred due to a variety of reasons. First, some mis-classified images are blurred or have bad lighting conditions, as shown in the examples in Figure 5(d), (h) and Figure 6(c). Secondly, some mis-classified images such as Figure 5(c) and 5(g) may contain

TABLE III
BINARY CLASSIFICATION OF INFECTION AND ISCHEMIA

(a) The performance measures of binary classification of infection by EfficientNet and baseline models

Model	Accuracy	F1-Score	AUC	Sensitivity	Specificity	TPR	FPR	Testing Time
EfficientNet B5	**0.9792**	**0.9792**	**0.9792**	**0.99**	0.97	**0.99**	0.03	**11m 17s**
EfficientNet B3	0.978	0.978	0.978	0.98	**0.98**	0.98	0.02	16m 9s
EfficientNet B2	0.9741	0.9741	0.9741	0.97	0.97	0.97	0.02	44m 50s
EfficientNet B4	0.9738	0.9738	0.9738	0.97	0.98	0.967	0.019	14m 32s
EfficientNet B1	0.9714	0.9714	0.9714	0.97	0.98	0.967	0.02	31m 60s
EfficientNet B0	0.968	0.968	0.968	0.96	0.97	0.96	0.03	45m 9s
EfficientNet B5 (no data augmentation)	0.9164	0.92	0.92	0.90	0.9333	0.90	0.0667	8m 6s
DenseNet121	0.943	0.943	0.943	0.95	0.94	0.9508	0.065	1h 17m 33s
ResNet50	0.87	0.87	0.87	0.85	0.89	0.85	**0.11**	59m 15s
Inception v3	0.86	0.86	0.86	0.84	0.88	0.84	0.12	1h 14m 47s
VGG16	0.85	0.85	0.85	0.83	0.87	0.83	0.13	1h 03m 22s
Ensemble CNN [12]	0.727	0.722	0.731	0.709	0.744			
3 Layer CNN	0.85	0.85	0.85	0.87	0.84	0.87	0.16	1h 04m 23s

(b) The performance measures of binary classification of ischemia by EfficientNet and Baseline models

Model	Accuracy	F1-Score	AUC	Sensitivity	Specificity	TPR	FPR	Testing Time
EfficientNet B1	**0.9939**	**0.9939**	**0.9939**	**1.00**	**0.99**	**1.00**	0.0123	4m 23s
EfficientNet B4	0.9919	0.9919	0.9919	0.99	0.99	0.9947	0.011	4m 23s
EfficientNet B0	0.9905	0.9905	0.9905	0.99	0.99	0.9918	**0.0107**	**4m 15s**
EfficientNet B5	0.9899	0.9899	0.9899	1.00	0.98	0.9986	0.018	4m 15s
EfficientNet B3	0.9892	0.9892	0.9892	1.00	0.98	1.00	0.02	5m 10s
EfficientNet B2	0.9885	0.9885	0.9885	1.00	0.98	0.9986	0.02	7m 44s
EfficientNet B1 (no data augmentation)	0.9946	0.99	0.99	1.00	0.9893	1.00	0.0107	10m 54s
Inception v3	0.9858	0.9858	0.9858	1.00	0.97	1.00	0.027	59m 15s
DenseNet121	0.9689	0.9689	0.9689	0.96	0.98	0.96	0.02	59m 15s
VGG16	0.9534	0.9534	0.9534	0.92	0.98	0.924	0.016	59m 15s
ResNet50	0.9467	0.9467	0.9467	0.91	0.98	0.9134	0.021	59m 15s
Ensemble CNN [12]	0.903	0.902	0.904	0.886	0.921			
3 Layer CNN	0.9413	0.9413	0.9413	0.98	0.90	0.979	0.1	59m 15s

ambiguous features that may be difficult to analyze even for wound experts.

Explainability using GradCam activation maps: Gradient-weighted Class Activation Mapping (or GradCam) [39] uses the gradients in the final convolutional layer of the target to produce a localization map that highlights regions in the image that were important for predicting the target. Four examples of wound images from the Infection dataset and their GradCam images are shown in Figure 6(a). Four examples of wound images from the ischemia dataset and their GradCam images are shown in Figure 6(b). The ischemia and non ischemia wound labels are shown under each column of the wound images.

In the GradCam maps, the red area highlights the region(s) the model focused on. The GradCam of DFU wound images show that for the diagnosis of infection and ischaemia, the EfficientNet model focused on the wound and the nearby skin with less focus on the corners of the images (blue and green areas). There is a blue and green area in the center of the second GradCam of ischaemia, which is counterintuitive. This might be because this wound was covered with necrotic tissue (dead tissue) and does not show the the actual wound. Consequently, the EfficientNet model focused on the nearby skin to discover visual cues of ischaemia, which is caused by insufficient blood that may affect the visual appearance of nearby skin.

5-fold cross validation boxplots: Figure 7 shows boxplots of the test accuracies of four experiments using 5-fold cross validation. The 5-fold cross validation experiment was performed with the EfficientNet B5 model on the Infection dataset and EfficientNet B1 model on the ischemia dataset with and without data augmentation. It can be observed that the variance in the EfficientNet model's accuracies are small (or high stability) across different folds as evidenced by the relatively small height of the boxplots for both the infection and ischemia datasets.

IV. DISCUSSION

Summary of findings: The EfficientNet architecture outperformed all baselines including Ensemble CNN, the prior state-of-the-art for both infection and ischemia classification across a comprehensive set of metrics including Accuracy, F1-Score, AUC, Sensitivity, Specificity, TPR and FPR. Its results were stable across 5 folds as evidenced by a small range in the boxplots across folds and results were generally better for the classification of ischemia than infection. Various versions of the EfficientNets were explored, showing that EfficientNet B5 performed best for infection classification, while EfficientNet B1 performed best for ischemia classification. Data augmentation improved the performance of EfficientNet B5 on the infection dataset but did not improve the performance of EfficientNet B1 on the ischemia dataset, possibly because the performance numbers were already high for ischemia classification. Due to

(a) Examples of non-infection cases correctly classified by the EfficientNet

(b) Examples of infection cases correctly classified by EfficientNet

(c) Examples of non-infection cases mis-classified by EfficientNet (both have ambiguous features)

(d) Examples of infection cases mis-classified by EfficientNet (blurred and bad lighting; ambiguous features)
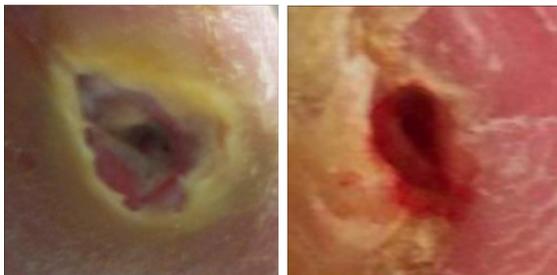
(e) Examples of non-ischemia cases correctly classified by EfficientNet models

(f) Examples of ischemia cases correctly classified by EfficientNet

(g) Examples of non-ischemia cases mis-classified by EfficientNet (both have ambiguous features)

(h) Examples of ischemia cases mis-classified by EfficientNet (ambiguous features; blurred and bad lighting)

Fig. 5. Examples of correctly and incorrectly classified images for infection and ischemia.

compound scaling, the test set speeds of EfficientNets were significantly faster than all baselines, which is important for the envisioned remote assessment scenario.

The EfficientNet model mis-classified some images for several reasons including those that were blurred, had bad lighting condition, or contained ambiguous and misleading features that may be challenging to analyze even for wound experts.

Limitations of this work: First of all, the datasets included cases of infection and ischemia that were visible and debrided

(removal of dead skin). In several cases, infection and ischemia may occur but may not always be visible or have to be analyzed without debridement. Secondly, although both conditions were analyzed separately, infection and ischemia can occur simultaneously on the same wound. Consequently, it might be meaningful to create a dataset that includes images with multiple conditions occurring in the same DFU image and investigate multi-label classification. Finally, the wound images were captured professionally with great care. However, in our envisioned use case, the
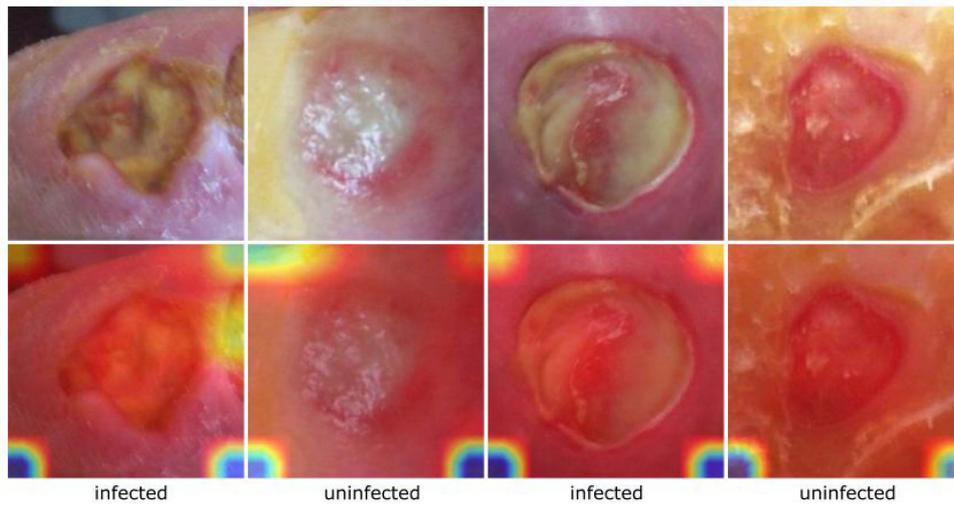
(a) wound images and their GradCam images by EfficientNet B5 on Infection dataset
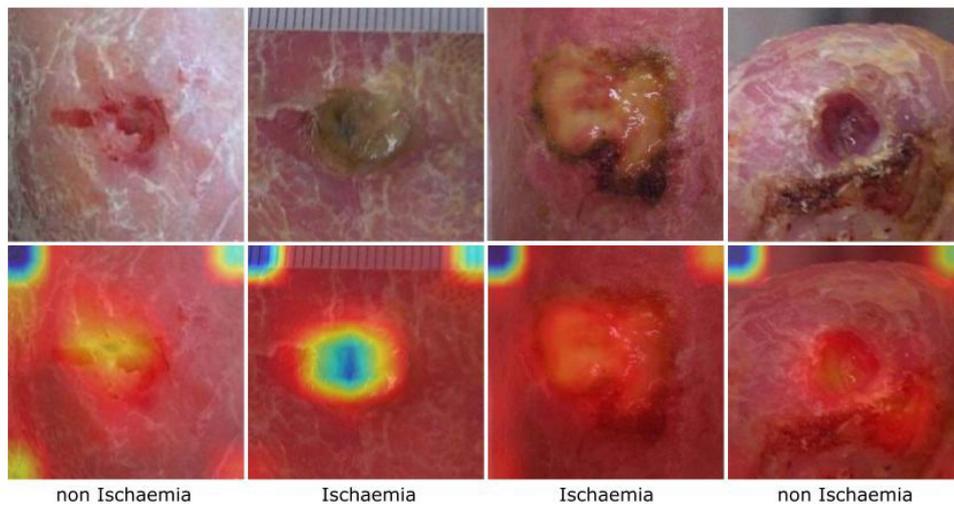
(b) wound images and their GradCam images by EfficientNet B1 on ischemia dataset

(c) images misclassified by the EfficientNet model; first row: Infection dataset; second row: ischemia dataset

Fig. 6. Examples of wound images GradCam and misclassified images.
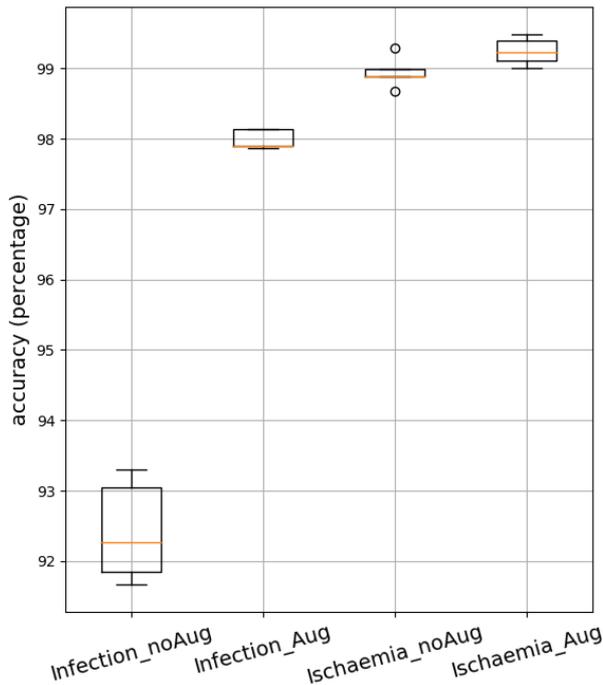
Fig. 7. boxplot for 5 fold cross validation using EfficientNet B5 on Infection dataset and EfficientNet B1 on ischemia dataset.

wound images will be captured by a wound nurse who will likely be a novice photographer. Consequently, many wounds may not occur in the center of the wound image, or maybe blurry or have arbitrary capture angles. The inclusion of such imperfect images would make the model's performance more robust and more realistic.

V. CONCLUSION

This work investigated using EfficientNet model for the binary classification of: (1) ischemia and non-ischemia; and (2) infection and non-infection on a Diabetic Foot Ulcers (DFU) dataset. We also evaluated the performance of our proposed models comprehensively and compared them to the results of other baseline models. EfficientNet models achieved excellent performance for the binary classification of ischemia and infection and outperformed all baselines including EnsembleCNN, the prior state-of-the-art for ischemia and infection classification, with a slightly higher accuracy in the classification of ischemia. Due to compound scaling, the EfficientNet was also significantly faster than all baselines. The envisioned wound diagnosis system that will integrate the proposed Infection and Ischaemia classification module could replicate the diagnoses rubric of the wound experts, supporting the work of non-experts in low resource settings and situations (e.g. patients' homes) where wound experts are not available.

The research described in this paper is related to our prior research on using a patch attention CNN with context preserving attention to predict the Photographic Wound Assessment Tool (PWAT) wound assessment score of chronic wounds [40]. The PWAT is considered the state-of-the-art rubric for scoring a wound's healing status based on key visual attributes including

wound size, depth, amounts of various tissues types, and periulcer viability. The Infection and Ischaemia classifiers developed in this work will be a separate module that flags wounds that need to be referred to experts and fast-tracked to prevent limb amputations. In future work, we would like to validate the proposed model in a live deployment.

REFERENCES

[1] K. Järbrink et al., "The humanistic and economic burden of chronic wounds: A protocol for a systematic review," *Systematic Rev.*, vol. 6, no. 1, pp. 1–7, 2017.
[2] S. R. Nussbaum et al., "An economic evaluation of the impact, cost, and medicare policy implications of chronic nonhealing wounds," *Value Health*, vol. 21, no. 1, pp. 27–32, 2018.
[3] L. Gould et al., "Chronic wound repair and healing in older adults: Current status and future research," *Wound Repair Regeneration*, vol. 23, no. 1, pp. 1–13, 2015.
[4] C. K. Sen et al., "Human skin wounds: A major and snowballing threat to public health and the economy," *Wound Repair Regeneration*, vol. 17, no. 6, pp. 763–771, 2009.
[5] M. A. Fonder, G. S. Lazarus, D. A. Cowan, B. Aronson-Cook, A. R. Kohli, and A. J. Mamelak, "Treating the chronic wound: A practical approach to the care of nonhealing wounds and wound care dressings," *J. Amer. Acad. Dermatol.*, vol. 58, no. 2, pp. 185–206, 2008.
[6] J. L. Mills Sr. et al., "The society for vascular surgery lower extremity threatened limb classification system: Risk stratification based on wound, ischemia, and foot infection (wifi)," *J. Vasc. Surg.*, vol. 59, no. 1, pp. 220–234, 2014.
[7] S. Beyaz, Ü. Ö. Güler, and G. Ş. Bağır, "Factors affecting lifespan following below-knee amputation in diabetic patients," *Acta Orthopaedica et Traumatologica Turcica*, vol. 51, no. 5, pp. 393–397, 2017.
[8] J.-L. Richard, A. Sotto, and J.-P. Lavigne, "New insights in diabetic foot infection," *World J. Diabetes*, vol. 2, no. 2, pp. 24–32, 2011.
[9] T. Gerassimidis, C. Karkos, D. Karamanos, and A. Kamparoudis, "Current endovascular management of the ischaemic diabetic foot," *Hippokratia*, vol. 12, no. 2, pp. 67–73, 2008.
[10] S. J. Landis, "Chronic wound infection and antimicrobial use," *Adv. Skin Wound Care*, vol. 21, no. 11, pp. 531–540, 2008.
[11] J. B. Rice, U. Desai, A. K. G. Cummings, H. G. Birnbaum, M. Skornicki, and N. B. Parsons, "Burden of diabetic foot ulcers for medicare and private insurers," *Diabetes Care*, vol. 37, no. 3, pp. 651–658, 2014.
[12] M. Goyal, N. D. Reeves, S. Rajbhandari, N. Ahmad, C. Wang, and M. H. Yap, "Recognition of ischaemia and infection in diabetic foot ulcers: Dataset and techniques," *Comput. Biol. Med.*, vol. 117, 2020, Art. no. 103616.
[13] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.
[14] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.
[15] J. J. van Netten, D. Clark, P. A. Lazzarini, M. Janda, and L. F. Reed, "The validity and reliability of remote diabetic foot ulcer assessment using mobile phone images," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, 2017.
[16] M. H. Yap, B. Cassidy, J. M. Pappachan, C. O'Shea, D. Gillespie, and N. D. Reeves, "Analysis towards classification of infection and ischaemia of diabetic foot ulcers," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform.*, 2021, pp. 1–4.
[17] M. H. Yap, C. Kendrick, N. D. Reeves, M. Goyal, J. M. Pappachan, and B. Cassidy, "Development of diabetic foot ulcer datasets: An overview," in *Diabetic Foot Ulcers Grand Challenge*. Berlin, Germany: Springer, 2021, pp. 1–18.
[18] Q. V. L. Mingxing Tan, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, vol. 97, pp. 6105–6114.
[19] M. Tan and Q. Le, "EfficientNetv2: Smaller models and faster training," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10096–10106.
[20] X. Zhao et al., "Fine-grained diabetic wound depth and granulation tissue amount assessment using bilinear convolutional neural network," *IEEE Access*, vol. 7, pp. 179151–179162, 2019.
[21] J.-T. Hsu, T.-W. Ho, H.-F. Shih, C.-C. Chang, F. Lai, and J.-M. Wu, "Automatic wound infection interpretation for postoperative wound image," *Proc. SPIE*, vol. 10225, 2017, Art. no. 1022526.

[22] J.-T. Hsu et al., "Chronic wound assessment and infection detection method," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, pp. 1–20, 2019.

[23] C. Wang et al., "A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks," in *Proc. IEEE 37th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2015, pp. 2415–2418.

[24] H. Nejati et al., "Fine-grained wound tissue analysis using deep neural network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 1010–1014.

[25] M. Goyal, N. D. Reeves, A. K. Davison, S. Rajbhandari, J. Spragg, and M. H. Yap, "DFUNet: Convolutional neural networks for diabetic foot ulcer classification," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 4, no. 5, pp. 728–739, Oct. 2020.

[26] N. Al-Garaawi, R. Ebsim, A. F. Alharan, and M. H. Yap, "Diabetic foot ulcer classification using mapped binary patterns and convolutional neural networks," *Comput. Biol. Med.*, vol. 140, 2022, Art. no. 105055.

[27] I. S. A. Abdelhalim, M. F. Mohamed, and Y. B. Mahdy, "Data augmentation for skin lesion using self-attention based progressive generative adversarial network," *Expert Syst. Appl.*, vol. 165, 2021, Art. no. 113922.

[28] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefer, "Skin lesion classification using ensembles of multi-resolution efficientnets with meta data," *MethodsX*, vol. 7, 2020, Art. no. 100864.

[29] F. Perez, C. Vasconcelos, S. Avila, and E. Valle, "Data augmentation for skin lesion analysis," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Cham, Switzerland: Springer, 2018, pp. 303–311.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[31] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2818–2826.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[33] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[34] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[35] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 4510–4520.

[36] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[37] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," 2017, *arXiv:1710.05941*.

[38] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.

[39] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 618–626.

[40] Z. Liu, E. Agu, P. Pedersen, C. Lindsay, B. Tulu, and D. Strong, "Comprehensive assessment of fine-grained wound images using a patch-based CNN with context-preserving attention," *IEEE Open J. Eng. Med. Biol.*, vol. 2, pp. 224–234, 2021.