IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 August 2023, accepted 2 September 2023, date of publication 6 September 2023,
date of current version 13 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3312531

RESEARCH ARTICLE

DFU-SIAM a Novel Diabetic Foot Ulcer Classification With Deep Learning

MOHAMMUD SHAAD ALLY TOOFANEE[1,2], SABEENA DOWLUT[2], MOHAMED HAMROUN[1,4], KARIM TAMINE[1], VINCENT PETIT[2], ANH KIET DUONG[3], AND DAMIEN SAUVERON[1]

[1]Department of Computer Science, XLIM, UMR CNRS 7252, University of Limoges, 87060 Limoges, France
[2]Applied Computer Science Department, Université des Mascareignes, Roches Brunes, Beau Bassin-Rose Hill, Mauritius
[3]Faculty of Science and Technology, University of Limoges, 87060 Limoges, France
[4]3iL Ingénieurs, 87015 Limoges, France

Corresponding author: Mohammud Shaad Ally Toofanee (stoofanee@udm.ac.mu)

This work was supported by the XLIM Laboratory of University of Limoges.

ABSTRACT Diabetes affects roughly 537 million people in the world, and it is predicted to reach 783 million by 2045. Diabetic Foot Ulcer (DFU) is a major issue with diabetes that may lead to lower limb amputation. The rapid evolution of DFU demands immediate intervention to prevent the terrible consequences of amputation and related complications.This research introduces a novel approach utilizing deep neural networks and machine learning for the accurate classification of diabetic foot ulcer (DFU) images. The proposed method harnesses the cutting-edge capabilities of Convolutional Neural Networks (CNN) and Vision Image Transformers (ViT) within a Siamese Neural Network (SNN) Architecture. By employing similarity learning, the model efficiently categorizes DFU images into four distinct classes: None, Infection, Ischemia, or Both. The training process involves the use of the DFU2021 dataset, with all ethical clearances duly obtained. Notably, the model exhibits remarkable performance on both the validation and test data, indicating a significant breakthrough in the field of DFU disease image classification. The potential of this innovative model extends beyond classification; it holds promise as an integral component of a comprehensive detection tool and longitudinal treatment protocol validation for DFU disease.

INDEX TERMS DFU, deep learning, CNN, vision transformers, Siamese network, similarity detection, DFU classification.

I. INTRODUCTION

Diabetes is one of the major diseases that affect roughly 537 million people, with a prediction to reach 783 million by 2045 [1]. Diabetic Foot Ulcer (DFU) is a prevalent complication among individuals with diabetes mellitus. In a recent research, it was found that mortality from DFU was high, with global mortality from diabetic foot ulcers standing at approximate 50% within 5 years [2]. This condition increases the risk of lower limb amputations in individuals with diabetes. When treating DFUs, promptness and assertiveness can make a significant difference in slowing the wound's course and preventing the need for an amputation [3].

Healing DFU can become a challenging and daunting task. Hence, to reduce the risk of DFU, certain preventive measures can be taken, such as 1) identification of the feet at risk 2) regular examinations. 3) raising awareness in the general public 4) treating risk factors. While it is important to design an early detection intervention, the integration of technology can also be explored to increase the accuracy of the results and the ease of examination procedures [4]. According to the Foot Care Clinic of APSA International, located in Mauritius, approximately 500 individuals undergo amputations each year as a result of type 2 diabetes, with an estimated 88% of these cases being preventable [5]. Annually, it is estimated that 67% of amputations in the United States and 90% of amputations in the United Kingdom are attributed to diabetes [6].

A. ARTIFICIAL INTELLIGENCE (AI) AND HEALTHCARE

Medical imaging data is one of the best sources of information about patients and helps to see inside the person's

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

body non-invasively or as least invasively as possible. In the current context of soaring demand for medical imaging and the prevailing challenge of staffing shortages in hospitals, the integration of AI tools holds promise as a potential solution [7]. The integration of digital imaging and AI has had a significant impact on the medical field. The use of AI algorithms has enhanced the accuracy and efficiency of these diagnoses. Research has shown that AI can accurately diagnose conditions such as retinal disorders [8], diabetic retinopathy [9], breast cancer [10], and skin cancer [11] by relying on digital images.

B. AI, HEALTHCARE AND ETHICS

While the use of AI in the field of healthcare is extremely promising and is progressing at a sustained pace, there are major challenges in terms of ethics and privacy [12]. Collecting healthcare data in a particular country that is stored on a remote server in another country with a different jurisdiction is one of the challenges of AI [13]. The governance of these health data, ensuring thoughtful aggregation and appropriate access to fuel innovation and improve patient outcomes and healthcare system efficiency while protecting the privacy and security of data subjects [14]. However, the World Health Organisation (WHO) acknowledged the potential of AI to enhance diagnosis, treatment, health research, and drug development, as well as support governments in carrying out public health functions such as surveillance and outbreak response, but also proposed strict guidelines that need to be followed [15] among which ethical guidelines are as follows:

- Avoid harming others.
- Promote the well-being of others
- Ensure that all persons are treated fairly.
- Deal with people in ways that respect their interests.

The above ethical considerations were taken into consideration while working on this research.

C. AIMS AND OBJECTIVES

DFUs are currently assessed by diabetes physicians and podiatrists in foot clinics and hospitals. There have been several research projects based on the application of AI and deep learning for DFU classification and even detection. The work is inspired by the successful implementation of AI in the medical field and takes advantage of the DFU Grand Challenge [16] which provides a labeled dataset with 4 classes that can be used by AI researchers to experiment and test the best model for classifications of DFU.

The DFU2021 Challenge dataset was obtained from the organisers of the DFU challenge. The objective was to experiment with Convolutional Neural Network coupled with latest vision-based transformers model which has achieved state-of-the-art performance on a number of computer vision benchmark including ImageNet classification and COCO object detection and instance segmentation [17]. This paper implements the use of Siamese Neural Network coupled with the K-nearest neighbors algorithm to implement the classification of images of DFU disease. The Siamese Neural Network architecture was first introduced in the early 1990s for the purpose of solving signature verification as an image matching problem [18].

Our research primarily aims to introduce an improved DFU classification model that outperforms the latest advancements in this field. This marks the initial phase in creating a tool for monitoring Diabetic Foot Ulcer treatment protocols initiated by medical practitioners. For successful integration into such protocols, the model's DFU classification must be accurate. The subsequent step involves using this model as an additional resource for healthcare professionals to validate their treatment approaches. However, we anticipate various challenges before reaching this milestone. These include addressing the limited data accessible to AI researchers and handling the sensitivity of healthcare data, particularly when involving images of body parts.

The structure of the rest of this paper is organized as follows: We first review some important concepts relating in machine learning in Section II. We then expose the related works III on machine learning models used in the medical field, and more specifically, machine learning applied to DFU classification. Section IV gives a detailed presentation of the proposed architecture of the DFU-SIAM. We then give a comprehensive presentation of experimental results and discussions in Section V. We finally conclude and explore further research opportunities in Section VI.

II. BACKGROUND AND PRELIMINARIES

Prior to taking a deep dive into the use of AI with regard to Diabetic Foot Ulcer we present some important background information on the techniques we will be using in this research.

A. ARTIFICIAL NEURAL NETWORKS (ANN)

ANN is a supervised learning algorithm that is inspired by the structure and functioning of the human brain [19]. It consists of interconnected nodes, as shown in Figure 1, which are also known as neurons, organized into layers. Each neuron processes input data and passes its output to the next layer, ultimately producing an output. ANNs are widely used for tasks such as pattern recognition, classification, and regression.

ANN consists of numerous neurons, also known as perceptrons, which are organised into layers. The layers communicate through the network's parameters, which are shown as arrows. These parameters encompass weights and biases. The weights control the importance of each input, while biases determine how easily a neuron fires or activates. The outcome generated during forward propagation, known as the predicted value, is matched against the corresponding actual value (ground truth) to evaluate the neural network's effectiveness. This evaluation is facilitated by using the loss function. The loss function is also referred to as the cost function. The aim is to minimize the loss. In the first iteration, the predicted values are far from the ground truth values,

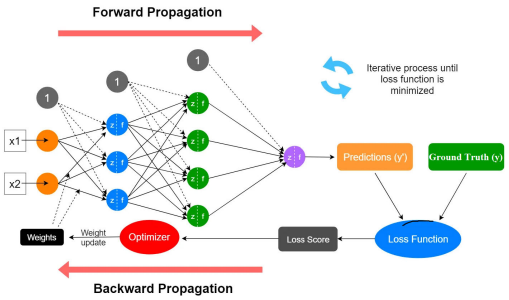


FIGURE 1. Graphical representation of an Artificial Neural Network showing the whole learning process, which consists of mainly three steps, which are: 1. Forward propagation; 2. Calculation of the loss function; and 3. Backward propagation [20].

and the loss will be high as weights and biases were initially assigned arbitrary values. The latter need to be updated in order to minimize the loss function, and this process of updating network parameters is called parameter learning or optimization which is done using an optimization algorithm (optimizer) that implements backpropagation. Backpropagation is a key step in training a neural network. It entails using the error from forward propagation to adjust the weights by propagating this loss backward through the network's layers This goes on for the number of epochs, which is fixed as a hyperparameter. An epoch is an iteration over the entire training dataset, which means the network has considered all the inputs once. An activation function is a function that is applied to the output of a neuron in an artificial neural network to determine whether the neuron will be activated or not. This is an important function as it introduces non-linearity in the network. There are several types of activation functions; among the most widely used are the softmax and sigmoid functions.

1) CONVOLUTIONAL NEURAL NETWORKS (CNN)

A CNN is a deep learning algorithm that utilizes convolutional operations to identify patterns within data, specifically image and video data. The convolution operation involves the application of a filter to the input data, through a process of sliding the filter over the data and computing the dot product between the filter and input. This produces a feature map, which summarizes the presence of distinct features within the input data. It was introduced by Lecun et al. [21] and has since achieved state-of-the-art performance in image classification task. CNN leverages the fact that nearby pixels are more strongly related than distant ones. It uses a special technique called convolution. Figure 2 shows a CNN that takes in an input image, assigns importance, learnable weights, and biases to various aspects of the image. The convolutional layers, explained in Figure 3 extract features from input data that are subjected to filters. This produces feature maps, which are passed into further processing layers.

The Max Pooling layer applies a pooling operation, which involves sliding a two-dimensional filter over each channel of

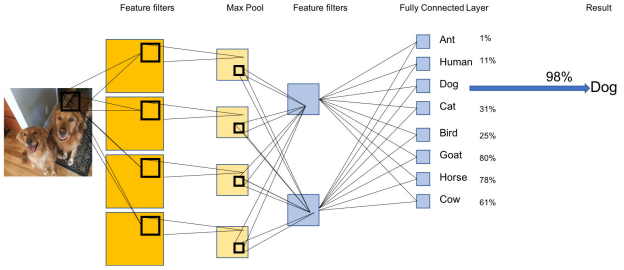


FIGURE 2. Simplified graphical representation of an CNN showing the processing of classification of images, including the convolutional layer, Max Pooling, and the Fully Connected layer [22].

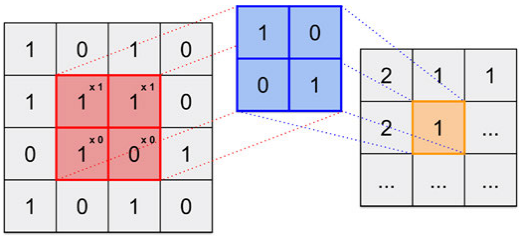


FIGURE 3. Illustration of a digital image undergoing convolution with a filter. The image on the left undergoes transformation into a feature map on the right, achieved through the use of a convolutional filter at the center. This specific filter is tailored to detect diagonal lines extending from the top left to the bottom right of the image. As the convolutional filter traverses the image in a predetermined manner, each element in the image (highlighted in red) is multiplied by its corresponding element in the convolutional filter (shown in blue). The sum of these products (depicted in orange) is then generated as output in a new matrix that indicates the presence of a diagonal line. In this feature map, a value of 2 indicates a complete diagonal line is detected, 1 suggests a portion of it is identified, and 0 signifies none of it is detected [23].

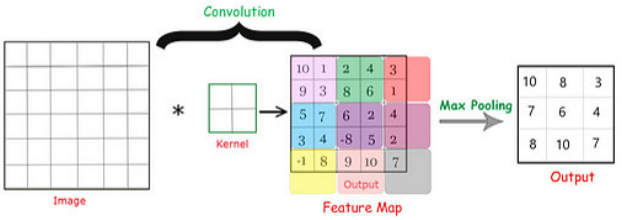


FIGURE 4. Illustration of Max Pooling in a CNN Network [24].

the feature map. The pooling layer summarizes the features lying in the region covered by the filter, thereby reducing the dimensions of the feature maps. This decreases the number of parameters to learn and the amount of computation performed in the network. There are different types of pooling layers, including Max pooling, average pooling, and global pooling. Figure 4 show ab example of applying Max pooling.

2) SIAMESE NEURAL NETWORK SNN

The Siamese network was presented in the context of signature verification [18] and comprises two identical networks that take in separate inputs, but are connected in the last layer. Siamese architecture aims to model semantic relationships between classes to extract discriminating features [25].

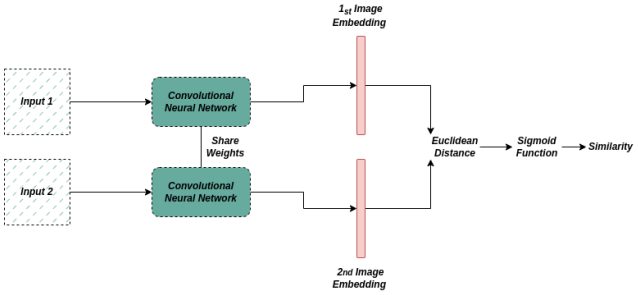


FIGURE 5. Illustration of a SSN with each sub-network consisting of a CNN architecture. The final output of the Siamese Network is a similarity score, which indicates how similar or different the two input images are. This score can be used to make a prediction, such as whether two images belong to the same class or not [27].

The twin networks have identical architecture, as shown in Figure 5. They also share weights and work in parallel to create vector representations for the inputs. For instance, we can use ResNet as the twin network if our inputs are images. This parallel CNN architecture allows the model to learn similarity, which can be used for tasks such as similarity measurement or classification. We can think of Siamese neural networks as wrappers for twin networks. They help produce better vector representations by measuring similarities between vectors. The loss function used by the SNN is the contrastive loss [26] which aims to maximise the proximity between positive pairs while simultaneously increasing the dissimilarity between negative pairs. Contrastive loss introduces the concept of margin, which is a minimal distance that dissimilar points need to keep. So it penalizes dissimilar samples for being closer than the given margin.

3) VISION IMAGE TRANSFORMERS (ViT)

Transformers are neural network architectures that have had a ground-breaking influence in the field of Natural Language Processing (NLP). Transformer architectures were able to tackle the shortcomings of sequential data tasks that were otherwise processed with Recurrent Neural Networks (RNNs). It came to the forefront in the famous paper "Attention Is All You Need" and uses self-attention mechanisms to capture the context of words in a sentence [28]. ViT was first introduced by the paper "An Image is Worth 16 × 16 Words: Transformers for Image Recognition at Scale" [29]. In ViTs, images are converted into sequences, enabling the models to predict class labels independently and learn image structures effectively. Each input image is treated as a sequence of patches, with each patch flattened into a single vector by concatenating the channels of its pixels. The resulting vectors are then linearly projected to achieve the desired input dimension. This approach allows ViTs to process images as sequences and capture their important features for classification tasks. This is illustrated in Figure 6.

Embedding is a process used in natural language processing for converting raw text into mathematical vectors since computers understand only 0s and 1s. As in any sentence, the exact position where a word is situated in a sentence can alter its meaning, so it is important to take this information into consideration. Positional embedding is a technique used in transformer models to add information about the position of a token in a sequence to its embedding.

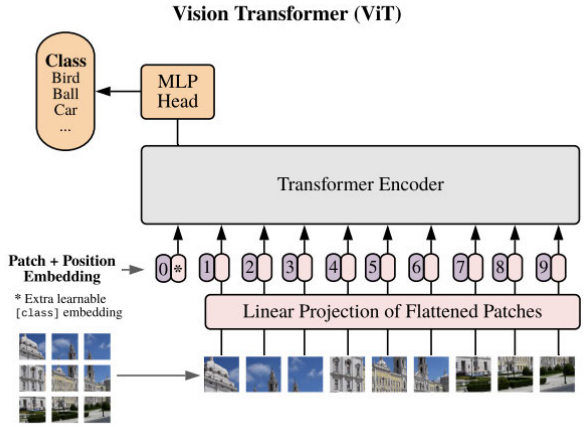


FIGURE 6. ViT splits an image into fixed-size patches, linearly embeds each of them, adds position embeddings, and feeds the resulting sequence of vectors to a standard Transformer encoder. In order to perform classification, it uses the standard approach of adding an extra learnable classification token to the sequence [29].

B. MACHINE LEARNING PROCESS WORKFLOW

Before explaining the proposed system, we shall explain the workflow when engaging in a machine learning research and development project.

1) Historical Data: In supervised learning, labeled data is necessary to acquire the ability to understand features. The initial phase involves ensuring the presence of dependable data. This process encompasses data collection and pre-processing to structure the data appropriately for supervised learning. In summary, we need data that has been accurately labeled and will be used for training the future machine learning model.
2) Model building: Depending on the desired objective, construct the model by meticulously selecting a suitable machine learning algorithm adapted for the problem you are looking to solve. Problems may include, for example, regression problems, classification problems, segmentation problems, and detection problems.
3) Model Evaluation: In order to enhance the potential for the algorithm to perform well on new, unseen data, it's common practice to divide the training dataset into a slightly smaller training set and a distinct validation set. The choice of evaluation metrics for assessing a model varies based on the nature of the model and whether it's being trained or tested. The validation set is designed to resemble the test dataset, aiding data scientists in refining an algorithm by pinpointing instances where the model could potentially generalize effectively and function within a novel population [23].
4) Model Optimisation: Machine learning algorithms consist of parameters. This stage focuses on utilizing the

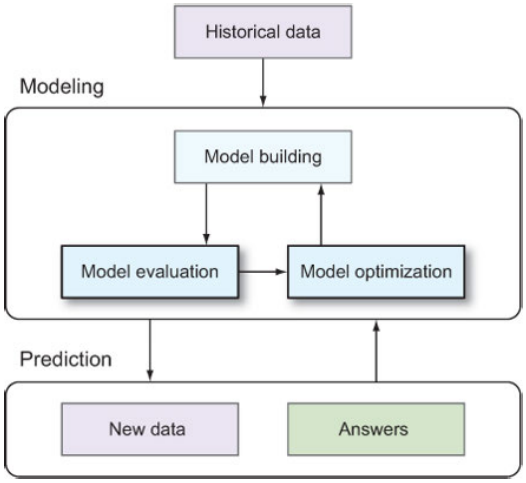


FIGURE 7. Workflow of real-world machine learning systems. From historical input data, you can build a model using a machine learning algorithm. You then need to evaluate the performance of the model and optimize its accuracy and scalability to meet your requirements. With the final model, you can make predictions on new data [30].

most suitable parameters to attain the highest level of performance.

III. RELATED WORKS

The potential AI has garnered attention from researchers who are aware of the risks posed by DFUs. In this section, we first study some recent work in the medical field and the AI algorithms used before taking a deep dive into our investigation of the use of AI in addressing the challenges posed by DFU.

A. AI TECHNIQUES AND HEALTHCARE

Electronic Health Record (EHR) systems store data associated with each patient encounter, including demographic information, diagnoses, laboratory tests and results, prescriptions, radiological images, and clinical notes [31]. These data are the raw materials used by researchers in AI and Healthcare. In this section, we look at the various AI algorithms that are used when dealing with this kind of data. Markovic et al. [32] applied a Novel NLP to medical records to create a patient-specific drug profile for diabetic patients. The degree of similarity between clustered profiles was calculated using Euclidean Distance. Yu et al. [33] also exploited transformers-based NLP Models to study social and behavioural determinants of health in Lung cancer atients. The models they used were BERT and RoBERTa, which are pre-trained language models. Furthermore, Soni et la. [34] also takes advantage of the use of Transformers, namely Bidirectional Encoder Representations from Transformers (BERT) for cohort retrieval in EHR.

Medical images have also been widely exploited by AI specialists in research. In their latest work, Yesilkaya et al. [35] tested nine classifiers, namely, nine distinct classifiers: k-nearest neighbors, decision tree, support vector machines, stochastic gradient descent, random forest, multi-layer perceptron, Naive Bayes, logistic regression, and AdaBoos for the detection of ovarian cancer. They apply feature reduction techniques to simplify their classification algorithms. In recent research, Rabiei et al. [36] tested random forest, neural network, gradient boosting trees, and genetic algorithms to mammographic features for prediction of breast cancer. Sajjad et al. [37] a pre-trained CNN model, VGG-19, is fine-tuned using augmented data for brain tumour grade classification into four classes.

Appiahene et al. [38] use non-invasive palm images to detect iron deficiency anaemia. They tested CNN, K-NN, Naive Bayes, Support Vector Machine, and Decision Tree after augmenting the dataset from 527 to 2635 images. In this case Naive Bayes achieved a 99.96% accuracy, while the Support Vector Machine achieved the lowest accuracy of 96.34%, and the CNN also performed better with an accuracy of 99.92% in detecting anaemia. In another non-invasive method, Al-Karawi et al. [39] experimented with severity-level detection of diabetic retinopathy using an ensemble model of CNN consisting of EfficientNetB7, ResNet50, and VGG19. The The classification accuracy achieved using the concatenation ensemble is 96%, which is higher than that obtained via individual CNN models. The use of ensemble techniques is promising and should be explored further.

B. AI TECHNIQUES AND DFU

In this section, we provide an in-depth review related to the main objective of this paper, which is the classification of DFU images using AI techniques.

The study by Galdran et al. [40] compares the performance of CNNs and ViTs [29] for the classification of DFUs. The authors investigated the efficacy of the ResNeXt50 [41] architecture from Big Image Transfer (BiT) [42] and EfficientNet [43] for CNNs, as well as the ViT and Data-efficient Image Transformers (DeiT) [44]. In addition, they compared the optimization approaches of Stochastic Gradient Descent [45] and Sharpness-Aware Optimization (SAM) [46] for neural network training. The authors employed various data augmentation techniques during training, including random rotations, horizontal/vertical flipping, and contrast/saturation/brightness adjustments. For testing, four versions of each image were generated, and the predictions were averaged for improved accuracy. Based on the results, the authors found that all pre-trained models performed better with the SAM optimizer. Specifically, the ResNeXt50 architecture demonstrated the highest performance on the test data. Interestingly, the authors achieved the highest scores by combining predictions from both CNN architectures. Through a thorough analysis of the various models, it can be concluded that CNNs outperform ViTs for the task of DFU classification.

Bloch et al. [47] introduced a novel approach for DFU classification using an ensemble of EfficientNets combined with

a semi-supervised training strategy incorporating pseudo-labeling [48]. They address the challenge of class imbalance in the dataset, by using Conditional Generative Adversarial Networks (GANs) [49] to generate synthetic DFU images. They utilized the pix2pixHD [50] framework for conditional image generation. The proposed pipeline consisted of three phases: baseline training, dataset extension, and extension training. In the baseline training phase, the best-performing models were combined into an ensemble model. The baseline model was then employed to train the GAN with pseudo-labeling for both unlabeled and test images. The resulting dataset was subsequently used to retrain the EfficientNet variants, and the best-performing models were merged for ensemble prediction. Notably, the proposed approach demonstrated improved performance of 55.80% compared to the work of Gladran et al. [40] 52.82% for ischaemia class F1 score.

In the study conducted by Ahsan et al. [51], the authors focused on investigating various CNN-based deep learning architectures for binary classification. Specifically, they evaluated the performance of AlexNet, VGG16/19, GoogLeNet, ResNet50.101, MobileNet, SqueezeNet, and DenseNet. Employing a fine-tuning approach, the authors conducted experiments and assessed the accuracy of each architecture. Notably, the results revealed that ResNet50 exhibited the highest accuracy among all the tested architectures. It is important to note that although this research is recent, it lacked comprehensive information regarding the hyperparameters used.

In their work, Goyal et al. introduced an ensemble CNN model that leverages the power of Inception-V3, ResNet50, and InceptionResNetV2 architectures [52], [53], [54]. The model combines bottleneck features extracted from these CNNs. During the training phase, the authors employed a strategy where the weights of the initial layers in the pre-trained networks were frozen to capture common features such as edges and curves. Subsequently, the later layers were unfrozen to focus on learning dataset-specific features. The ensemble-CNN model used the combined bottleneck features as input for binary classification, employing the Support Vector Machine algorithm. Comparative analyses were conducted against traditional machine learning methods, including BayesNet, Random Forest, and CNN-only approaches. Notably, the proposed CNN ensemble model outperformed all traditional machine learning techniques and CNN-only models, showcasing its superior performance in binary classification tasks.

Santos et al. [55] presented DFU-VGG, an innovative approach for the classification of diabetic foot ulcers (DFUs). The authors employed the VGG-19 architecture as the backbone of their CNN. Notably, they introduced batch normalization after each convolutional block. The performance of DFU-VGG was evaluated against fine-tuned versions of VGG-16, VGG-19, InceptionV3, ResNet50, DenseNet201, MobileNetV2, and EfficientNetB0 networks in their original configurations. In a separate study conducted by Santos et al. [56], an experiment was conducted to investigate the performance of an ensemble model comprising various combinations of VGG-16, VGG-19, InceptionV3, ResNet50, and DenseNet201 architectures. The outcome of the experiment showed that the ensemble model consisting of VGG-16, VGG-19, and DenseNet201 demonstrated the highest performance among the tested combinations. This research sheds light on the effectiveness of ensemble models in improving the classification performance of DFU classification. Thotad et al. [57], also proposed to use a fine-tuned CNN backbone based on EfficientNet [43].

Khandakar et al. [58] approach for DFU classification by combining a CNN-based backbone with traditional machine learning algorithms. The features of DFU are extracted using a pre-trained CNN model. Then unsupervised method of k-mean clustering is used for clustering the images into three categories, namely mild, moderate, and severe.

Qayyum et al. [59] experimented with CNN and ViT. They used transformer-based architectures that were originally trained on the ImageNet dataset. The different vision transformers are fine-tuned by adding a fully connected layer with feature size (3072×768), ReLU activation (ReLU), dropout layer for regularization, and another fully connected layer with feature size (768×4) at the end layer of the different pre-trained transformers. The features extracted from the last layer of multiple transformers are concatenated pair-wise and applied to a fully connected layer at the end to concatenate the features of individual transformers and then pass to the classifier layer. Table 1 provides a concise overview of the outcomes from the research related to machine learning and DFU classifications.

Based on the findings from above, it is evident that ensemble methods have demonstrated favorable outcomes, as has the use of CNN architectures for image-related tasks. Additionally, promising results have been observed with the application of transformer-based architectures for image classification. In light of this, our research will use these insights and introduce an innovative model for DFU classification, which will be elaborated further in the forthcoming section IV.

IV. PROPOSED ARCHITECTURE

This research proposes an innovative architecture for the classification of images of DFU diseases, as illustrated in Figure 8. We use an ensemble of CNN (Section II-A1) and ViT (Section II-A3) as the backbone for the two ''identical twins'' networks of the SNN (Section II-A2) for feature extraction. A detailed explanation of the model training process can be found in Section IV-C. The training procedure involves utilizing DFU images and employing k-fold validation with K set to 5. A comprehensive description of the dataset used and pre-processing applied in this study is provided in Section V-A. Once the model is trained, it becomes proficient in classifying input images into one of the four classes: none, infection, ischaemia, or both. To generate predictions on the test images, a k-Nearest Neighbors

TABLE 1. Summary of related works.

Research Work	Architecture	Type of Classifcation
Thotad et al. [57]	EfficientNet	Binary Class Abnormal / Normal
F Santos et al. [55]	VGG-16, VGG-19, InceptionV3, ResNet50, DenseNet201, MobileNetV2 EfficientNetB0	Multi-Class None / Infection/ Ischaemia / Both
E Santos et al. [56]	CNN Ensemble [VGG-16, VGG-19, InceptionV3, ResNet50, DenseNet201, MobileNetV2]	Multi-Class None / Infection/ Ischaemia / Both
Galdran et al. [40]	CNN [BIT- ResNeXt50, EfficientNet] ViT[ViT-base ,DeiT-small]	Multi-Class None / Infection/ Ischaemia / Both
Qayyum et al. [59]	ViT[vit_base_patch16_224]	Multi-Class None / Infection/ Ischaemia / Both
Ahmed et al. [60]	EfficientNet B0-B6 Resnet-50	Multi-Class None / Infection/ Ischaemia / Both
Bloch et al. [47]	EfficientNets B0,B1,B2 Pseudo-Labeling GAN	Multi-Class None / Infection/ Ischaemia / Both
Ahsan et al. [51]	AlexNet, VGG16/19, GoogLeNet, ResNet50.101, MobileNet, SqueezeNet, and DenseNet	Binary Class Infection / schaemia
Khandakar et al. [58]	K-Mean Clustering CNN	Multi-Class Mild / Moderate / High
Goyal et al. [61]	Ensemble CNN Support Vector Machine	Binary Class Infection / Ischaemia

(kNN) model is integrated into the approach, conducting neighbourhood analysis to enhance the prediction of the model. A detailed explanation of this process is presented in Section IV-C.

The dataset used in this work was obtained from the DFUC2021 challenge [62], as detailed in Section V-A. Due to the imbalanced nature of the dataset, various image augmentation techniques were employed to enhance the training process of the Siamese model. For both training and classification tasks, the KNN classifier was utilized.

A. DFU-SIAM

DFU-SIAM is a DFU disease classification model that implements a SNN. Figure 9 shows the ensemble model architecture. For the CNN backbone, we use EfficientNetV2S based on EfficientNet [43] architectures, which have been shown to significantly outperform other networks in classification tasks while having fewer parameters. EfficientNetV2S has fewer parameters, making it more suitable for low-resource settings, and it uses a combination of efficient network design and compound scaling to achieve high accuracy with fewer parameters [63].

The second backbone of the ensemble model is based on ViTs (Section II-A3), more specifically, Bidirectional Encoder representation from Image Transformers (BEiT). BEiT uses a pre-training task called masked image modeling (MIM) and stands for Bidirectional Encoder representation from Image Transformers, which draws inspiration from BERT [64]. MIM uses two views for each image, namely, image patches and visual tokens. The image is split into a grid of patches that are the input representation of the backbone Transformer. The image is ''tokenized'' into discrete visual tokens. During pre-training, some proportion of image patches are randomly masked, and the corrupted input is fed to Transformer. The model learns to recover the visual tokens of the original image instead of the raw pixels of masked patches.

The vector representation of image 1 is passed into both the EfficientNet model and the ViT model. In the EfficientNet, we remove the last dense layer from the pre-trained model to obtain the features from the last flattened layer (average pool). In the ViT model, we obtain the last hidden states, which contain all the patches from the last attention layer, except the classification token; then we flatten them and use another dense to reduce the shape to make the output

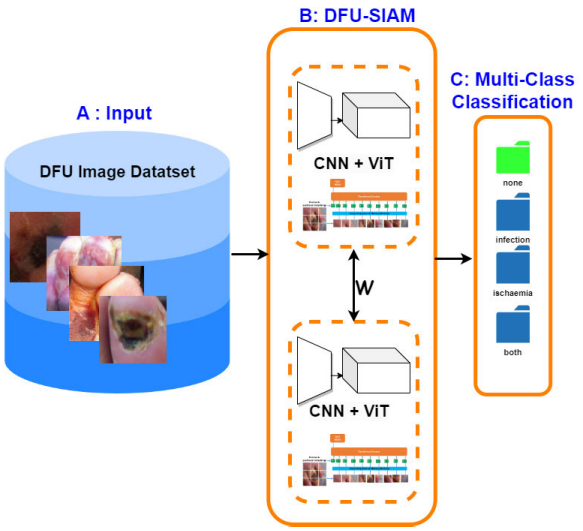


FIGURE 8. DFU-SIAM Architecture Overview for DFU Classification. A: Input images were sourced from the DFU2021 Dataset used for initial training and validation. B: The proposed Network, consisting of an ensemble of CNN and ViT within a Siamese Architecture. C: Visualization of the four distinct classes into which the DFU images are accurately classified.

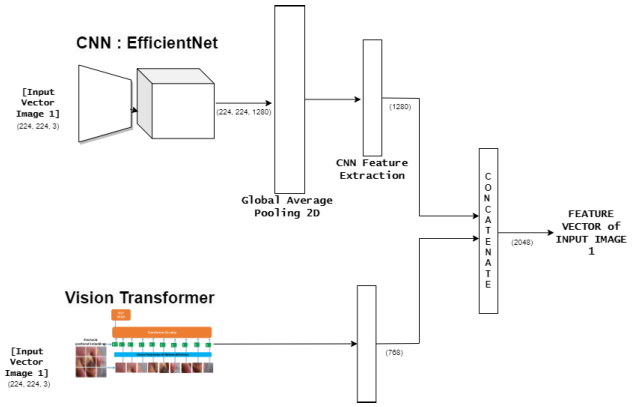


FIGURE 9. Block Diagram of the Ensemble Network, illustrating the internal architecture of the individual networks composing the SNN. The CNN utilized is EfficientNet, while the ViT employed is BEiT.

(features) have the same size as the feature extracted from EfficientNet. Finally, we merge the two feature sets.

Traditional Artificial Neural Networks learn by trying to minimise the loss function. Siamese Neural Network uses a different loss function, which is explained in the next section.

B. LOSS FUNCTION OF DFU-SIAM

While Siamese networks normally use contrastive loss, for DFU-SIAM we chose to implement Large Margin Cotangent Loss (LMCoT). Duong et al. [65] proposed LMCot as a novel approach for enhancing performance in verification and identification tasks. The LMCot loss utilizes the cotangent function instead of the cosine function. The cotangent function has a broader range of values, allowing for better optimization. Experimental results demonstrated that LMCot outperformed

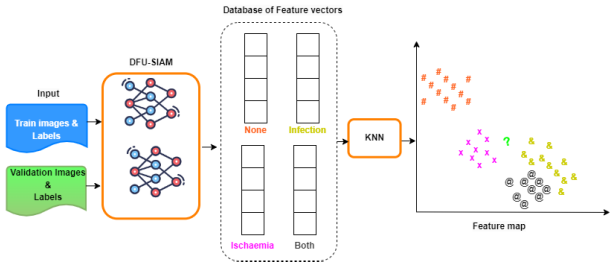


FIGURE 10. Illustration of the training process of DFU-SIAM, demonstrating the integrated approach of utilizing the SNN for feature extraction and machine learning for prediction during the training phase.

existing methods in various benchmark datasets and achieved state-of-the-art performance.

Once the chosen loss function has been established, it becomes crucial to outline the model evaluation and optimisation phase as detailed in Section II-B. In the subsequent section, we will provide a comprehensive explanation of how we intend to execute these steps to ensure optimal performance of the model.

C. DFU-SIAM MODEL EVALUATION AND OPTIMISATION

This section explains the training, validation, and prediction processes of DFU-SIAM. This process is mentioned in Section II-B where we explain the machine learning process workflow. The augmented dataset, consisting of training and validation images along with their respective labels, is loaded into the model. The model leverages the feature extraction capabilities of twin models to obtain the feature vectors of each image and employs the Large Margin Cotangent Loss as loss function. The objective of the learning process is to iteratively update the model parameters in order to minimize the distance between encoded features when the input images belong to similar classes while maximizing the distance when the input images belong to dissimilar classes. This ensures that the model learns to effectively discriminate between different classes by capturing meaningful patterns and representations in the encoded feature space.

During the validation and prediction processes of the model, it is important to mention that the KNN classifier [66], is used as depicted in Figure 11. We iterate through values of K from 1 to 30 to determine the optimal value of K. The metric we use to get the best K is Macro F1 score. KNN is a classifier model based on nearest neighborhood density estimation. For each epoch, an attempt is made to identify the optimal K value based on the Macro F1-score. Once the best K value is determined for an epoch, the corresponding weights are saved, and predictions are made on the test dataset. This iterative process ensures that the best predictions are obtained for each epoch of the test data.

The whole process of how DFU-SIAM classifies a test image is shown in Figure 12. The test image is fed into DFU-SIAM, which performs encoding and generates a compact feature vector within a lower-dimensional space. Within this

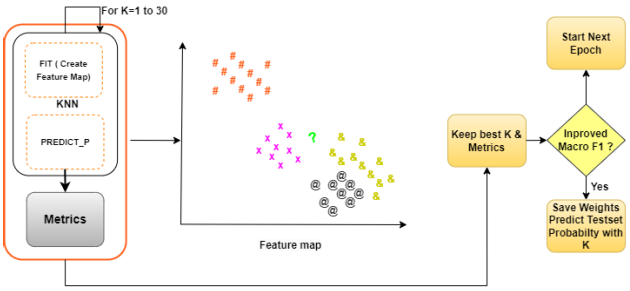


FIGURE 11. Schematic representation of the training- process of DFU-SIAM including making predictions using machine learning algorithm KNN to determine the optimal value for K based on the highest macro-F1 score The identified parameters are saved and subsequently employed for predictions on the test data.

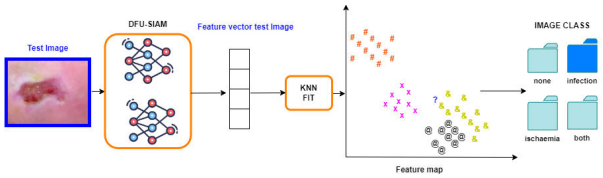


FIGURE 12. An overview of the application of DFU-SIAM for Test image classification. Input images are fed into DFU-SIAM, where they are encoded to generate feature vectors. The network then measures the distances between these feature vectors and all the training images. Utilizing the KNN algorithm, the predicted class for the input image is determined based on its proximity to the training samples.

reduced feature space, the encoded representation of the test image is compared to that of all the training samples using suitable distance measures. The classification is then carried out by employing the KNN algorithm.

V. EXPERIMENTATION AND RESULTS

This section starts by providing a detailed description of the dataset utilized for the experimentation, including details on the preprocessing techniques employed. Additionally, the materials used in the experiments are outlined, and the results obtained are presented alongside a thorough comparison with relevant works in the field.

The quality of the dataset significantly influences the performance of deep learning models in terms of result accuracy. However, ethical reliability of the data source is equally important. In the next section, we will define the characteristics of the dataset employed in DFU-SIAM.

A. DATASET

In this section, we give an overview of the dataset we will use for this research.

Data quality is a crucial factor that directly affects the performance of supervised learning algorithms. The utilization of a representative and high-quality dataset is critical for achieving optimal accuracy and performance [67]. In this study, we obtained the dataset from the DFUC2021 challenge organized by the Medical Image Computing and Computer-Assisted Intervention (MICCAI) society [62]. The proper licensing was also secured for this research, ensuring that all ethical and legal requirements were met.

Upon initial preprocessing, we observed that the dataset class distribution was imbalanced, with 621, 2555, 227, and 2552 instances belonging to the both, infection, ischaemia, and none categories, respectively, as shown in figure 13. Such an imbalance poses a challenge to the performance of supervised learning algorithms, as they tend to be biased towards the majority class. To address this issue, we applied data augmentation techniques, as discussed in Section V-A1. It should be noted that Siamese networks, when combined with data augmentation techniques, can enhance the performance of various tasks. Data augmentation introduces variations to the training data.

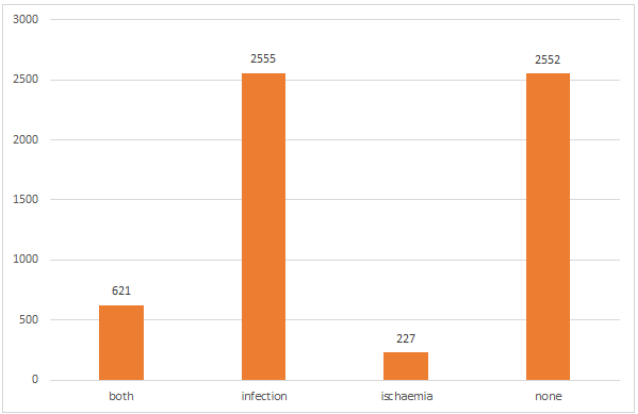


FIGURE 13. Class distribution of the DFU2021 Challenge dataset, illustrating the evident imbalance in the dataset.

1) DATA AUGMENTATION

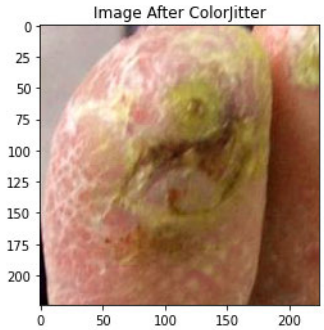
Imbalanced data refers to a situation where one class of data examples has much more representation than the other classes [68]. The geometric transformations that were applied to our DFU dataset set images are illustrated in figure 14 and include:

- Colorjitter (brightness = 0.1,contrast = 0.1, Saturation = 0.1, hue = 0.1) Figure 14a
- RandomEqualize(p = 0.2) Figure 14b
- RandomHorizontalFlip(p = 0.2) Figure 14c
- RandomVerticalFlip(p = 0.2 Figure 14d

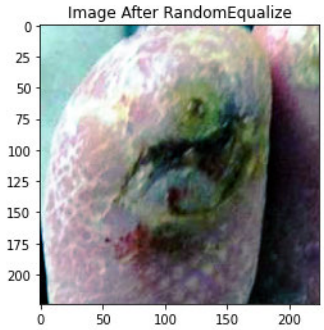
Prior to executing the model, it is most important to establish a suitable hardware and software setup, as they have an impact on the hyperparameters that will be employed. This setup is elaborated on below.

B. EXPERIMENTAL SETUP

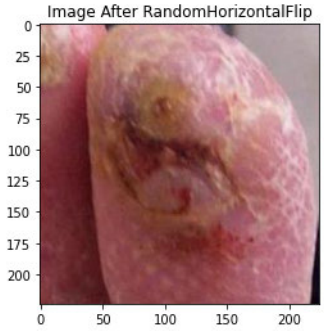
The experimental setup was conducted on a Windows 10 Pro operating system running on a powerful hardware configuration comprising 64 GB of RAM and an Intel(R) Xeon(R) W-2155 CPU operating at 3.30 GHz. The system was further enhanced with an NVIDIA GeForce RTX 3060 GPU, boasting 12 GB of dedicated memory. To facilitate the



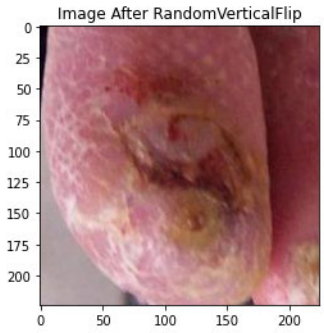
(a) Color Jitter



(b) Random Equalize



(c) Random Horizontal



(d) Random Vertical

FIGURE 14. Demonstration of the application of geometric image transformations (a) Color Jitter, (b)Random Equalize, (c) Random Horizontal Flip, (d) Random Vertical Flip.

experiments, the system was configured with CUDA version 11.7, Tensorflow 2.10.0, and Python 3.10.9.

The selection of hyperparameters in this study was influenced by the computational resources available. The batch size was set to 8, and the input images were resized to dimensions of 200 by 200 pixels with RGB channels. All models were run for 40 epochs. A fixed learning rate of 10e-6 was employed. To optimize the parameters for prediction on the test data, the KNN algorithm was utilized. Additionally, test time augmentation (TTA) [69] techniques were applied to further enhance the prediction accuracy. TTA introduces random modifications to the test images, enabling the trained model to encounter augmented versions of the images multiple times. The predictions for each corresponding image were averaged, providing a more robust and reliable final prediction.

As explained in the previous section, our intention is to employ an ensemble of CNN and ViT as identical sub-networks of the SNN. The next section will explain the backbone that will be used.

C. EXPERIMENTAL STRATEGY

For experimental strategy, we tested an ensemble of different combinations of CNN based and ViT based models. For the CNN, we maintained the EFFicientNet. However, for the ViT we experimented with BEiT [64], [70] and SwinTiny(SwinT) [71]. Both will be tested and evaluated against related works.

D. METRICS

In this study, we used various evaluation metrics to evaluate the performance of our classification algorithms. Commonly used metrics include precision, recall, and F-score, which are essential in quantifying the accuracy and effectiveness of a proposed method and also bench-marking it against other proposed models. This section briefly explains the metrics we used.

The selection among several configurations should be made without subjectivity. Therefore, the following sections will elaborate on the various machine learning metrics that will guide us in determining which model demonstrates the best performance.

1) CONFUSION MATRIX

A confusion matrix is an N X N matrix, where N is the number of classes being predicted. For the DFU problem at hand, we have N = 4, and hence we get a 4 × 4 matrix. Figure 15 shows an example of a confusion matrix for DFU classification.

2) PRECISION

The precision metric is determined by dividing the number of correctly classified positive samples by the total number of samples classified as positive, including those that were classified incorrectly. This metric serves as an indicator of the model's ability to accurately classify samples as positive. The formula is shown in 1 where TP refers to the True positive and FP represents the False positive.

$$Precision = \frac{TP}{TP + FP}. \quad (1)$$

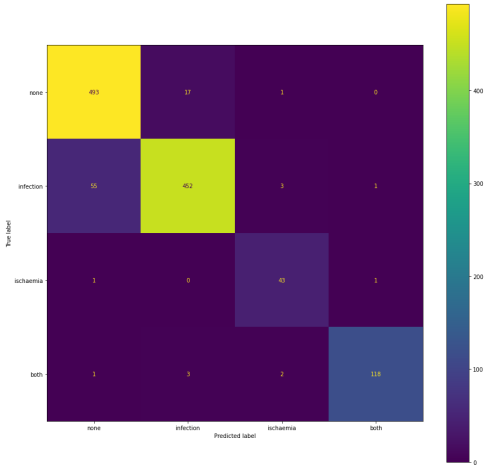


FIGURE 15. Example of a Confusion Matrix displaying the True labels and Predicted labels.

3) RECALL

Recall is derived by dividing the number of positive samples that were correctly classified as positive by the total number of positive samples in the dataset, as illustrated in the formula 2. This metric is used to evaluate the model's capacity to accurately identify positive samples. Higher values of recall indicate that the model is better at detecting positive samples.

$$Recall = \frac{TP}{TP + FN}. \tag{2}$$

4) F1-SCORE

F1-score is the harmonic mean of precision and recall values for a classification problem. F1 score symbolise high precision as well as high recall. It presents a good balance between precision and recall and gives good results on imbalanced classification problems. The following formula: 3

$$F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN}. \tag{3}$$

5) MACRO AVERAGE F1-SCORE

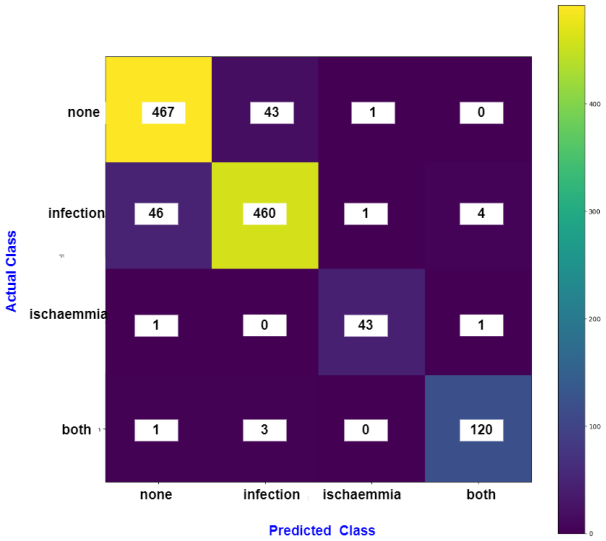
In multi-class classification with imbalanced data, the main consideration will be Macro F1-Score. The formula is illustrated with the following formula: 4 where n represents the number of classes involved. In the DFU classification. n is equal to 4.

$$Macro\ F1\ Score = \frac{\sum_{i=1}^{n} F1\ score}{n}. \tag{4}$$

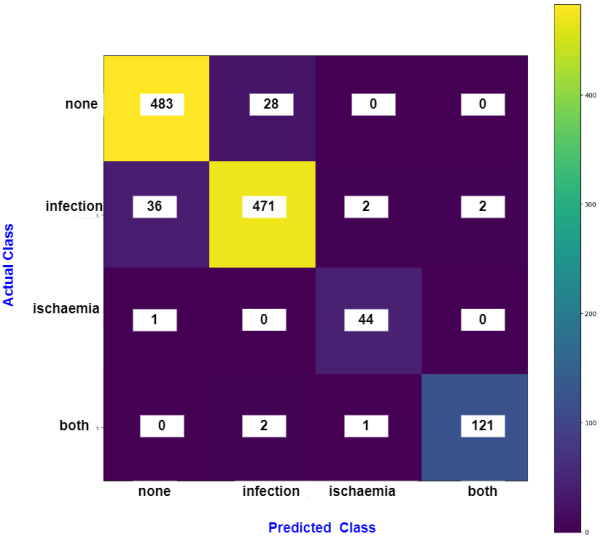
E. RESULTS

1) CONFUSION METRICS

The confusion matrix was obtained for the two variations of EfficientNet and ViT transformer, as shown in Figure 16. Figure 16a shows the confusion matrix when the backbone



(a) CM : EffNet-SwinT



(b) CM: EffNet-BEiT

FIGURE 16. Confusion matrix results obtained from applying two different ensembles as identical networks (a) Confusion Matrix with an ensemble of EfficientNet and SwinT. (b) Confusion Matrix with an ensemble of EfficientNet and BeiT.

of our model is run with EfficientNet as the CNN backbone and SwinT(EfficientNEt/SwinT) as the vision Transformer. Figure 16b shows the confusion matrix with EfficientNet and BeIT (EfficientNet/BEiT) as combined backbones. From the overall confusion matrix, the performance metrics are calculated. Table 2 and Table 3 show these metrics. By analysing the confusion matrices, we see that both models are wrongly predicting some instances of none class as infection and some as infection as none.

From Table 2 and Table 3 we can see that EfficientNet and BEiT has a better accuracy of 95% compared to 93% of EfficientNet and SwinT. The Macro F1-score is same at 0.95. EfficientNet and BEiT model has a higher macro F1-score for the classes none and infection.

TABLE 2. Metrics from confusion matrix EfficientNet/SwinT.

	Precision	Recall	F1-Score
none	0.92	94	0.93
infection	0.93	0.91	0.92
ischaemia	0.96	1	0.98
both	0.98	0.98	0.98
accuracy			0.93
macro avg.	0.95	0.96	0.95
weighted avg.	0.93	0.93	0.93

TABLE 3. Metrics from confusion matrix EfficientNet and BEiT.

	Precision	Recall	F1-Score
none	0.93	0.95	0.94
infection	0.94	0.92	0.93
ischaemia	0.94	0.98	0.96
both	0.98	0.98	0.98
accuracy			0.95
macro Avg.	0.95	0.96	0.95
weighted avg.	0.94	0.94	0.94

2) LOSS

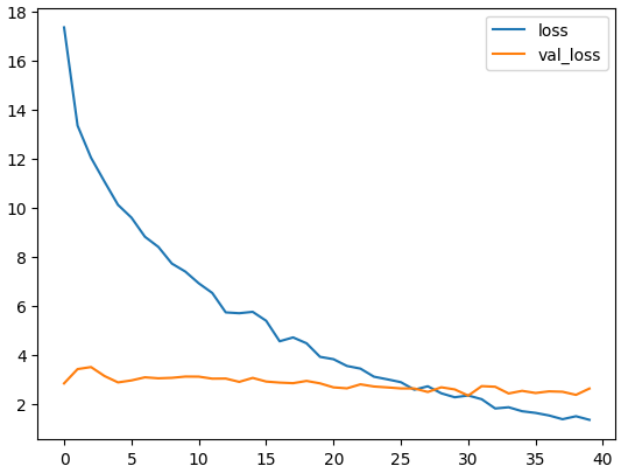
The loss function provides insights into the effectiveness of the models in minimizing errors and improving their predictive performance. By analyzing the loss curves, we can observe the behavior of the models over time and assess their training progress. As far as training loss is concerned, we can see in Figure 17 for both models that training loss decreases at a constant rate. This indicates effective learning and model improvement throughout the training process. Furthermore by analysing validation loss curves, we can assess how well the models are learning and how effectively they are adapting to the validation dataset. The validation loss very quickly stagnates for both models. However, we can witness a constant decrease for the EfficientNet and BEiT model as shown in Figure 17b.

3) ACCURACY

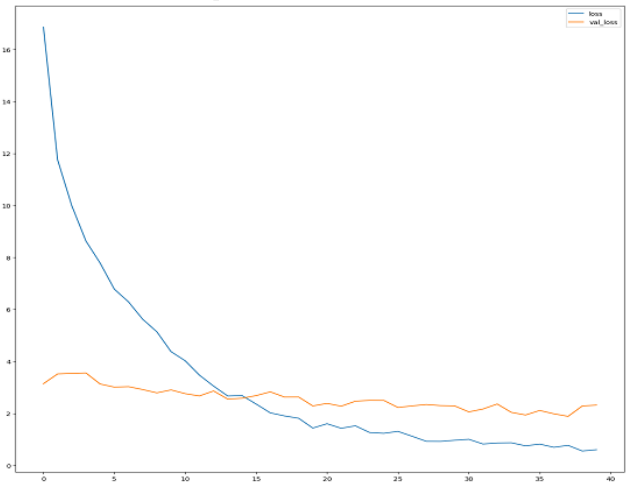
In the best case, for a deep learning model, we would like both curves to increase harmoniously during the training process, indicating that the model is learning and improving its performance on both the training and validation datasets. In Figure 18a training and learning curves intersect at around epoch 30 while in Figure 18b the intersection is earlier at around epoch 13. If the change continues to increase with validation and training accuracy diverging, this will signal that the model is overfitting. In the current case, while there seems to be a discrepancy, we do not believe that the model is overfitting. However, this shows that there is room to further investigate and improve performance.

4) MACRO F1-SCORE

In figure 19 we show how the Macro-F1 score varies during the 40 epochs. Figure 19a shows that a high Macro F1 score is obtained very early, at epoch 15. However, a look at



(a) Loss curve of the model with an ensemble of EfficientNet and SwinT, trained for 40 epochs.



(b) Loss curve of the model with an ensemble of EfficientNet and BEiT, trained for 40 epochs.

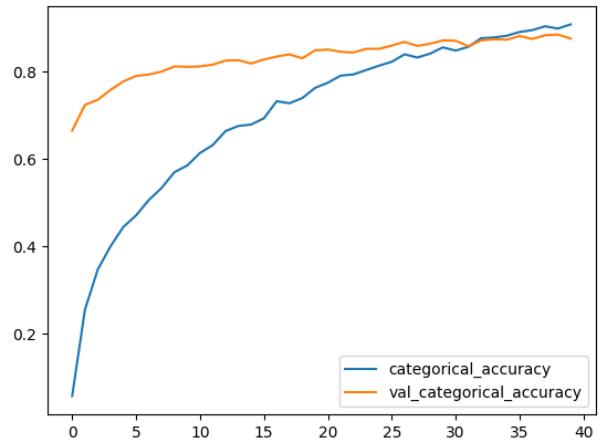
FIGURE 17. Loss curves of the two models being experimented.

Figure 19b shows a peak at epoch 17 but it has another peak at epoch 38. This indicated that it is a good idea to investigate both models on unseen test data to have a better indication of which is most suited for the DFU disease classification.

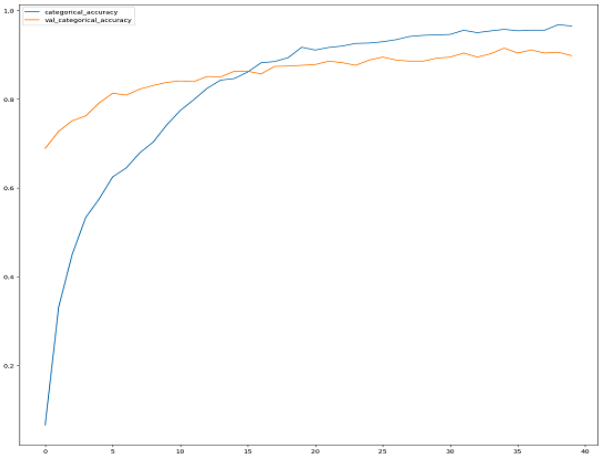
5) SUMMARY OF METRIC AND COMPARISON

In this section, a summary of the two models is shown in table 4. For this table it can be seen that based on the main metric on which we are evaluating our model the EfficientNet and SwinT model has a higher macro F1-score compared to the EfficientNet and BEiT model. The class F1-score for none,infection is better for EfficientNet and BEiT while for ischaemia EfficientNet and SwinT is better. For both class they have same class F1-score.

The two models were evaluated on test data provided by the DFU2021 Challenge. This consists of 5734 unlabeled images that are used to make predictions and uploaded on the platform to get the required metrics.



(a) Accuracy of the model with an ensemble of EfficientNet and SwinT, trained for 40 epochs

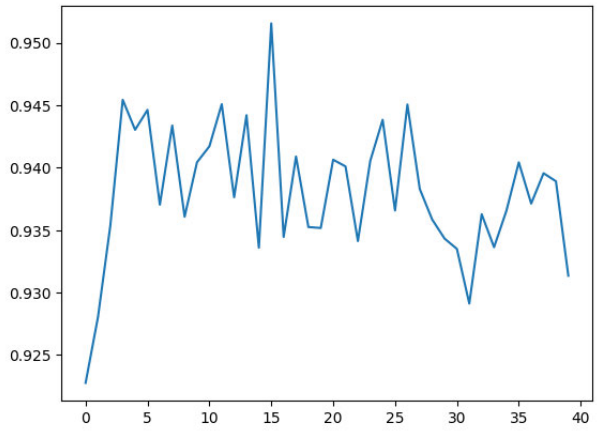


(b) Accuracy of the model with an ensemble of EfficientNet and BEiT, trained for 40 epochs

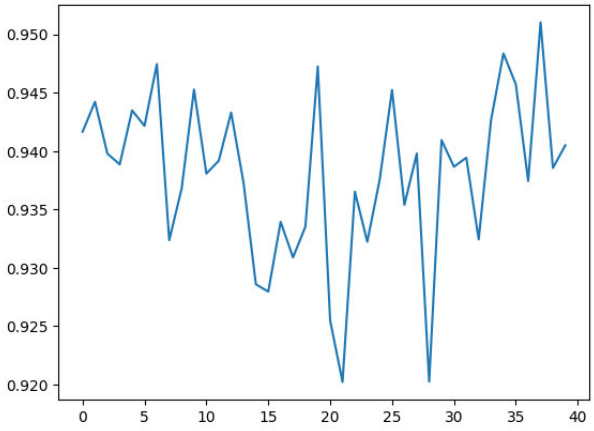
FIGURE 18. Accuracy curves of the two models being experimented.

TABLE 4. Comparison of CNN and ViT siamese models.

Model	EfficientNet and SwinT	EfficientNet and BeIT
Macro F1-Score	**0.9516**	0.9510
loss	5.3856	**0.6980**
categorical_ accuracy	0.6922	**0.9556**
val_loss	2.9078	**1.9856**
val_categorical _accuracy	0.8270	**0.9110**
Accuracy	0.9320	**0.9395**
None F1-Score	0.9265	**0.9370**
Infection F1-Score	0.9217	**0.9308**
Ischaemia F1-Score	**0.9783**	0.9565
Both F1-Score	0.9798	0.9798
Macro Precision	**0.9477**	0.9472
Macro Recall	**0.9558**	0.9551
Macro AUC	0.9748	**0.9781**
Weighted Avg. Precision	0.9322	**0.9397**
Weighted Avg. Recall	0.9320	**0.9395**
Micro F1-Score	0.9320	**0.9395**
Epoch #	**15**	38



(a) Macro F1 Score variation over the 40 epochs of training for the model with an ensemble of EfficientNet and SwinT, exhibiting a peak at epoch 15.



(b) Macro F1 Score variation over the 40 epochs of training for the model with an ensemble of EfficientNet and BEiT, exhibiting a peak at epoch 38

FIGURE 19. Macro F1-Score variation of the two models being experimented over 40 epochs.

TABLE 5. Performance on test data.

Metrics	EfficientNet/SWINT	EfficientNet/BEiT
Macro F1-Score	0.5850	**0.6160**
None F1-Score	0.7442	**0.7478**
Infection F1-Score	0.6072	**0.6149**
Ischaemia F1-Score	0.5367	**0.5613**
Both F1-Score	0.4520	**0.5401**
Macro Precision	0.5892	**0.6115**
Macro Recall	0.6368	**0.6570**
Macro AUC	0.8043	**0.8298**
Weighted Avg. Precision	**0.6818**	0.6728
Weighted Avg. Recall	0.6610	**0.6918**
epochs	16	**7**

Table 5 presents the performance of the two models on test data which were uploaded on the DFU2021 live challenge board. From the table, it can clearly be observed that the EfficientNet and BEiT model exhibits better overall performance in almost all the metrics except for the weighted average precision. Hence, the EfficientNet and BEiT model was further optimised. The predictions that showed the highest macro-F1 score were averaged and loaded on the classification liveboard.

TABLE 6. DFU-SIAM comparison with related works.

Metrics	DFU-SIAM	Galdran et al. [40]	Bloch et al. [47]	Ahmed et al. [60]	Qayyum et al. [59]
Rank	**BEST**	1st	2nd	3rd	4th
Macro F1-Score	**0.6228**	0.6216	0.6077	0.5959	0.5691
None F1-Score	0.7553	**0.7574**	0.7453	0.7157	0.7466
Infection F1-Score	0.6276	0.6388	0.5917	**0.6714**	0.6281
Ischaemia F1-Score	0.5495	0.5282	**0.558**	0.4574	0.467
Both F1-Score	**0.5588**	0.5619	0.5359	0.539	0.4347
Macro Precision	0.5486	0.614	**0.6207**	0.5984	0.5814
Macro Recall	**0.6554**	0.6522	0.6246	0.5979	0.6104
Macro AUC	0.8599	**0.8855**	0.8616	0.8644	0.8488
W.Avg. Precision	0.6983	**0.7009**	0.6853	0.6730	0.68
W.Avg Recall	0.6815	**0.6856**	0.6657	0.6711	0.6636
Micro F1-Score	0.6749	**0.6801**	0.6532	0.6714	0.6577
epochs	Avg(best epoch)	NA	NA	NA	5(ended)

When compared to the performance of related works, DFU-SIAM which is a model based on a siamese neural network for DFU disease classification, exhibits the best Macro F1-Score as shown in Table 6. Galdran et al. [40] (Galdran 22) were actually the winners of the DFU challenge.

F. DISCUSSION

DFU classification is implemented using a Siamese Neural Network which is in itself a novel architecture, combined with Large Margin Cotangent Loss (LMCot) as a novel approach for enhancing performance in verification and identification. We further introduce the KNN classifier while iteratively searching for the best K while doing prediction on test data. These are the reasons that explain why our model, DFI-SIAM, performs better than the other model in the related work. While Galdran et al. [40] focused on comparing Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) and achieved the best macro F1-score, our approach takes a different direction by combining these two architectures. By incorporating the strengths of both CNNs and ViTs, we capitalize on their complementary features and achieve improved results. As far as Bloch et al. [47] they used an ensemble of EfficientNet families with pseudo-labeling. In DFU-SIAM we choose EfficientNet, or more precisely, EfficientNetV2S, which is one of the best performing pre-trained CNN. Qayyum et al. [59] concentrated essentially on vision transformers. They propose the combination of two different pre-trained ViT models for feature extraction. For our proposed model, we chose BEiT, which is one of the best performing pre-trained transformers. However, we decided to make the last 10 layers of the BEiT transformer trainable as our experiments showed a significant increase in performance.

DFU classification in our study used a novel approach of using innovative SNN architecture for classification of DFU. To further enhance its performance, we chose to use a novel approach called the Large Margin Cotangent Loss (LMCot) proposed by Duoung et al. [65]. Our proposed model, DFI-SIAM, surpasses the performance of other models discussed in related works.

Bloch et al. [47] employed an ensemble of EfficientNet models with pseudo-labeling, which differs from our methodology. Instead, we specifically chose the EfficientNetV2S model, known for its outstanding performance as a pre-trained CNN. We acknowledge, however, that the pseudo-labeling can be used in our model to further improve its performance.

Furthermore Qayyum et al. [59] concentrated on Vision Transformers, proposing the combination of two distinct pre-trained ViT models for feature extraction. In our study, we adopt the BEiT model, which exhibits very good performance as a pre-trained transformer. However, we make a deliberate choice to train only the last 10 layers of the BEiT transformer to strike a balance between fine-tuning and computational efficiency.

By integrating these advancements and tailoring them to the specific requirements of DFU classification, the DFU-SIAM model achieves remarkable accuracy and sets a new benchmark in the field. It should be noted that the model's computational efficiency was not evaluated at this stage. This parameter is important if the model is to be deployed on ubiquitous devices. One limiting factor of the system is the imbalanced data, and this is also acknowledged by other researchers, with Bloch et al. [47] using pseudo-labeling and Generative Adversarial Network to tackle this.

While exploring need for more data, We may have clinics or medical centers that adhere to the idea of using deep learning models but are not willing to share the data with third parties. Ensuring patient privacy while integrating diverse datasets into a model has emerged as a significant limitation in deep learning research [72]. This problem can be addressed by deploying the model using Federated Learning [73]. The notable aspect of Federated Learning lies in its ability to handle data in a decentralized manner, thereby fostering a privacy-preserving environment in AI applications [74] in the

event we require several distant sites to contribute to having even more data, which is an important aspect for the training and implementation of a deep learning model.

G. LIMITATIONS

One limitation of our study arises from the substantial class imbalance present within the dataset, particularly evident in the under representation of ''both'' and the Ischaemia class class. Upon careful inspection of the images, we observed that certain geometric data augmentation techniques were already applied to these classes during the dataset creation process. This imbalance has influenced the overall performance of the models. Nevertheless, it is worth noting that the DFU2021 dataset is currently the most comprehensive resource available for conducting research in this domain. Furthermore, we remain optimistic that with adequate computational resources, there is potential to explore additional variations and employ ensemble modeling techniques to enhance the outcome of our study.

VI. CONCLUSION AND FUTURE WORKS

In this paper, we have trained and tested a new model based on an ensemble of EfficientNet and BeIT Transformer in a SNNmodel that has outperformed some of the best results obtained for classification of DFU as detailed in related works III. The dataset limitations can be addressed in future work by investigating the use of GAN which is a type of deep neural network that consists of two components: a generator network and a discriminator network [49]. Another option would be using pseudo labeling which is a technique used in machine learning to improve model performance by using unlabeled data in conjunction with labeled data [48]. The 5734 unlabeled data in the test image can thus be exploited.

This research marks an important step towards tackling the use of machine learning in the field of DFU image classification. Despite our limited processing power, we effectively utilized available resources to achieve significant results. With access to greater computational capabilities, we anticipate that further fine-tuning of our model will lead to even better performance.

As previously specified, there is a need to have a better-quality and more balanced dataset to curb data bias and ensure the model generalises well to unseen data. One possible solution that should be explored is accessing data collected at different geographically located medical facilities. This clearly poses the problem of data privacy, as the owner of the data would not want highly sensitive health-related data to be transferred to a third party. To overcome this barrier, the use of centralized Federated Learning or Peer-to-peer Federated Learning should be explored. Federated Learning, an innovative distributed interactive AI concept, holds exceptional promise in the realm of intelligent healthcare. This approach enables multiple clients, including entities like hospitals, to engage in AI training while upholding stringent data privacy protocols [75], [76], [77]. It entails the training of machine learning models across datasets dispersed throughout various data centers, such as hospitals and clinical research labs, all while safeguarding data integrity [78].

Incorporating data from a variety of sources will undoubtedly contribute to enhancing the dataset's imbalance, thereby alleviating the data bias observed in the ''both'' and ''ischaemia'' classes. These classes currently exhibit only 621 and 227 occurrences, in contrast to the ''infection'' and ''none'' classes which encompass 2555 and 2552 instances, respectively. It's important to highlight that ''both'' and ''ischaemia'', which are the most serious forms of DFU, are relatively less prevalent in the samples. However, this poses a challenge for machine learning algorithms. One potential approach to addressing this imbalance is to employ GAN [79] for generating synthetic images. This technique has been successfully employed by Kim et al. [80] to augment liver ultrasonic image data using a semi-supervised approach.

This work serves as a stepping stone for future research and development aimed at effectively detecting, treating, and managing diabetic foot ulcers. Our ultimate goal is to contribute to advancements in the medical field, leading to improved patient outcomes and healthcare management. As the machine learning model learns by trying to reduce the loss to a minimum, it is prone to making erroneous predictions. If data bias is present, then there will most certainly be errors in predictions. Hence, from a medical point-of-view, it is mandatory to explainability on top of clinicial validation [81]. The critical obstacle to the widespread acceptance of machine learning in healthcare and research relates to the black box nature of machine learning algorithms for the end user [82]. There is presently extensive research concentrating on Explainable AI (XAI), which aims to provide a suite of machine learning techniques that enable human users to understand, appropriately trust, and produce more explainable models [83]. This has to be given priority in any future work. One simple step could be to show a class activation mapping (CAM) approach that highlights the infected section or section with ischaemia and improves the visual interpretability [84]. Techniques like LIME (Local Interpretable Model-Agnostic Explanations) [85] and SHAP (Shapley Additive exPlanations) should be explored [86].

A crucial future direction for our research involves utilizing the Siamese Neural Network to develop a tool that can aid medical practitioners in evaluating the treatment protocols they administer to patients over time. This longitudinal disease evaluation tool would enable practitioners to monitor and adjust treatments as needed. Subsequently, after thorough testing and evaluation of the tool, it can be adapted into a preventive tool for early detection of DFU disease in patients, accessible via a mobile phone platform. In order to advance to the next phase, our plan involves collaborating with experts from public health research labs who possess the necessary expertise in designing protocols for assessing the effectiveness and acceptability of technology adoption in healthcare settings.

ACKNOWLEDGMENT

The authors would like to thank the Research Team U1094 Inserm U270 IRD EpiMaCT ''Épidémiologie des maladies chroniques en zone tropicale'' of the University of Limoges, which has contributed to this work by offering the processing power of their server for the GPU processing.

REFERENCES

[1] (2021). *IDF Diabetes Atlas, 10th EDN. International Diabetes Federation*. Brussels, Belgium. [Online]. Available: https://diabetesatlas.org/idfawp/resource-files/2021/07/IDF_Atlas_10th_Edition_2021.pdf

[2] L. Chen, S. Sun, Y. Gao, and X. Ran, ''Global mortality of diabetic foot ulcer: A systematic review and meta-analysis of observational studies,'' *Diabetes, Obesity Metabolism*, vol. 25, no. 1, pp. 36–45, 2023.

[3] R. G. Frykberg, ''Diabetic foot ulcers: Pathogenesis and management,'' *Amer. family physician*, vol. 66, no. 9, p. 1655, 2002.

[4] I. D. A. Rismayanti, V. N. Farida, N. W. S. Dewi, R. Utami, A. Aris, and N. L. P. I. B. Agustini, ''Early detection to prevent foot ulceration among type 2 diabetes mellitus patient: A multi-intervention review,'' *J. Public Health Res.*, vol. 11, no. 2, p. jphr.2022.2752, Apr. 2022.

[5] A International. (2021). *Foot Care*. [Online]. Available: https://apsa.mu/foot-care/

[6] N. Ahmad, G. N. Thomas, P. Gill, and F. Torella, ''The prevalence of major lower limb amputation in the diabetic and non-diabetic population of England 2003–2013,'' *Diabetes Vascular Disease Res.*, vol. 13, no. 5, pp. 348–353, Sep. 2016.

[7] C. McCague, K. MacKay, C. Welsh, A. Constantinou, R. Jena, and M. Crispin-Ortuzar, ''Position statement on clinical evaluation of imaging ai,'' *Lancet Digital Health*, vol. 5, no. 7, pp. 400–402, 2023.

[8] F. Li, H. Chen, Z. Liu, X. Zhang, and Z. Wu, ''Fully automated detection of retinal disorders by image-based deep learning,'' *Graefe's Arch. Clin. Experim. Ophthalmol.*, vol. 257, no. 3, pp. 495–505, Mar. 2019.

[9] R. Gargeya and T. Leng, ''Automated identification of diabetic retinopathy using deep learning,'' *Ophthalmology*, vol. 124, no. 7, pp. 962–969, Jul. 2017.

[10] A. Rodríguez-Ruiz, E. Krupinski, J.-J. Mordang, K. Schilling, S. H. Heywang-Köbrunner, I. Sechopoulos, and R. M. Mann, ''Detection of breast cancer with mammography: Effect of an artificial intelligence support system,'' *Radiology*, vol. 290, no. 2, pp. 305–314, Feb. 2019.

[11] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, ''Dermatologist-level classification of skin cancer with deep neural networks,'' *Nature*, vol. 542, pp. 115–118, Jan. 2017.

[12] I. Bartoletti, ''Ai in healthcare: Ethical and privacy challenges,'' in *Artificial Intelligence in Medicine*, D. Riano, S. Wilk, and A. ten Teije, Eds. Poznań, Poland: Springer, 2019, pp. 7–10.

[13] V. N. Shah and S. K. Garg, ''Managing diabetes in the digital age,'' *Clin. Diabetes Endocrinol.*, vol. 1, no. 1, pp. 1–7, Dec. 2015.

[14] J. S. Winter, ''AI in healthcare: Data governance challenges,'' *J. Hospital Manag. Health Policy*, vol. 5, p. 8, Mar. 2020.

[15] *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*, World Health Organization, Geneva, Switzerland, 2021.

[16] M. H. Yap, B. Cassidy, and C. Kendrick, *Diabetic Foot Ulcers Grand Challenge*. Berlin, Germany: Springer, 2022.

[17] Z. Xie, Y. Lin, Z. Yao, Z. Zhang, Q. Dai, Y. Cao, and H. Hu, ''Self-supervised learning with Swin transformers,'' 2021, *arXiv:2105.04553*.

[18] J. Bromley, I. Guyon, Y. LeCun, E. Säckinger, and R. Shah, ''Signature verification using a 'Siamese' time delay neural network,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 6, 1993.

[19] H. H. Aghdam and E. J. Heravi, *Guide to Convolutional Neural Networks*, vol. 10. New York, NY, USA: Springer, 2017, p. 51.

[20] R. Pramoditha. (2022). *Overview of a Neural Network's Learning Process*. Accessed: Aug. 17, 2023. [Online]. Available: https://medium.com/data-science-365/overview-of-a-neural-networks-learning-process-61690a502fa

[21] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, ''Gradient-based learning applied to document recognition,'' *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[22] P. Lea, *Internet of Things for Architects: Architecting IoT Solutions by Implementing Sensors, Communication Infrastructure, Edge Computing, Analytics, and Security*. Birmingham, U.K.: Packt Publishing, 2018.

[23] R. Y. Choi, A. S. Coyner, J. Kalpathy-Cramer, M. F. Chiang, and J. P. Campbell, ''Introduction to machine learning, neural networks, and deep learning,'' *Transl. Vis. Sci. Technol.*, vol. 9, no. 2, p. 14, 2020.

[24] K. Prasant. (2021). *Max Pooling, Why Use it and Its Advantages*. Accessed: Aug. 21, 2023. [Online]. Available: https://medium.com/geekculture/max-pooling-why-use-it-and-its-advantages-5807a0190459

[25] S. Berlemont, G. Lefebvre, S. Duffner, and C. Garcia, ''Class-balanced Siamese neural networks,'' *Neurocomputing*, vol. 273, pp. 47–56, Jan. 2018.

[26] R. Hadsell, S. Chopra, and Y. LeCun, ''Dimensionality reduction by learning an invariant mapping,'' in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2006, pp. 1735–1742.

[27] N. Georgios. (2023). *How Do Siamese Networks Work in Image Recognition? | Baeldung on Computer Science*. Accessed: Aug. 17, 2023. [Online]. Available: https://www.baeldung.com/cs/siamese-networks

[28] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, ''Attention is all you need,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[29] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, ''An image is worth 16×16 words: Transformers for image recognition at scale,'' 2020, *arXiv:2010.11929*.

[30] H. Brink, J. Richards, and M. Fetherolf. (2016). *Real-World Machine Learning*. Simon and Schuster. [Online]. Available: https://livebook.manning.com/book/real-world-machine-learning/chapter-1/

[31] G. S. Birkhead, M. Klompas, and N. R. Shah, ''Uses of electronic health records for public health surveillance to advance public health,'' *Annu. Rev. Public Health*, vol. 36, no. 1, pp. 345–359, Mar. 2015.

[32] R. Markovič, V. Grubelnik, T. Završnik, H. B. Vošner, P. Kokol, M. Perc, M. Marhl, M. Zavrsnik, and J. Zavrsnik, ''Profiling of patients with type 2 diabetes based on medication adherence data,'' *Frontiers Public Health*, vol. 11, Jul. 2023, Art. no. 1209809.

[33] Z. Yu, X. Yang, C. Dang, S. Wu, P. Adekkanattu, J. Pathak, T. J. George, W. R. Hogan, Y. Guo, and J. Bian, ''A study of social and behavioral determinants of health in lung cancer patients using transformers-based natural language processing models,'' in *Proc. AMIA Annu. Symp.*, 2021, p. 1225.

[34] S. Soni and K. Roberts, ''Patient cohort retrieval using transformer language models,'' in *Proc. AMIA Annu. Symp.*, 2020, p. 1150.

[35] B. Yesilkaya, M. Perc, and Y. Isler, ''Manifold learning methods for the diagnosis of ovarian cancer,'' *J. Comput. Sci.*, vol. 63, Sep. 2022, Art. no. 101775.

[36] R. Rabiei, ''Prediction of breast cancer using machine learning approaches,'' *J. Biomed. Phys. Eng.*, vol. 12, no. 3, p. 297, Jul. 2022.

[37] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. W. Baik, ''Multi-grade brain tumor classification using deep CNN with extensive data augmentation,'' *J. Comput. Sci.*, vol. 30, pp. 174–182, Jan. 2019.

[38] P. Appiahene, J. W. Asare, E. T. Donkoh, G. Dimauro, and R. Maglietta, ''Detection of iron deficiency anemia by medical images: A comparative study of machine learning algorithms,'' *BioData Mining*, vol. 16, no. 1, pp. 1–20, Jan. 2023.

[39] A. Al-Karawi and E. Avşar, ''A deep learning framework with edge computing for severity level detection of diabetic retinopathy,'' *Multimedia Tools Appl.*, pp. 1–22, Mar. 2023.

[40] A. Galdran, G. Carneiro, and M. A. G. Ballester, ''Convolutional nets versus vision transformers for diabetic foot ulcer classification,'' in *Diabetic Foot Ulcers Grand Challenge*. Cham, Switzerland: Springer, 2022, pp. 21–29.

[41] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, ''Aggregated residual transformations for deep neural networks,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1492–1500.

[42] A. Kolesnikov, L. Beyer, X. Zhai, J. Puigcerver, J. Yung, S. Gelly, and N. Houlsby, ''Big transfer (BiT): General visual representation learning,'' in *Computer Vision—ECCV*. Glasgow, U.K.: Springer, 2020, pp. 491–507.

[43] M. Tan and Q. Le, ''EfficientNet: Rethinking model scaling for convolutional neural networks,'' in *Proc. 36th Int. Conf. Mach. Learn.*, vol. 97, K. Chaudhuri and R. Salakhutdinov, Eds. 2019, pp. 6105–6114.

[44] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jegou, ''Training data-efficient image transformers & distillation through attention,'' in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 10347–10357.

[45] S. H. Haji and A. M. Abdulazeez, "Comparison of optimization techniques based on gradient descent algorithm: A review," *PalArch's J. Archaeol. Egypt/Egyptol.*, vol. 18, no. 4, pp. 2715–2743, 2021.

[46] P. Foret, A. Kleiner, H. Mobahi, and B. Neyshabur, "Sharpness-aware minimization for efficiently improving generalization," 2020, *arXiv:2010.01412*.

[47] L. Bloch, R. Brungel, and C. M. Friedrich, "Boosting efficientnets ensemble performance via pseudo-labels and synthetic images by pix2pixHD for infection and ischaemia classification in diabetic foot ulcers," in *Diabetic Foot Ulcers Grand Challenge*. Cham, Switzerland: Springer, 2022, pp. 30–49.

[48] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2013, vol. 3, no. 2, p. 896.

[49] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[50] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8798–8807.

[51] M. Ahsan, S. Naz, R. Ahmad, H. Ehsan, and A. Sikandar, "A deep learning approach for diabetic foot ulcer classification and recognition," *Information*, vol. 14, no. 1, p. 36, Jan. 2023.

[52] 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. (CVPR)*, Jun. 2016, pp. 2818–2826.

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

[54] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 4278–4284.

[55] F. Santos, E. Santos, L. H. Vogado, M. Ito, A. Bianchi, J. M. Tavares, and R. Veras, "DFU-VGG, a novel and improved VGG-19 network for diabetic foot ulcer classification," in *Proc. 29th Int. Conf. Syst., Signals Image Process. (IWSSIP)*, Jun. 2022, pp. 1–4.

[56] E. Santos, F. Santos, J. Dallyson, K. Aires, J. M. R. S. Tavares, and R. Veras, "Diabetic foot ulcers classification using a fine-tuned CNNs ensemble," in *Proc. IEEE 35th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jul. 2022, pp. 282–287.

[57] P. N. Thotad, G. R. Bharamagoudar, and B. S. Anami, "Diabetic foot ulcer detection using deep learning approaches," *Sensors Int.*, vol. 4, 2023, Art. no. 100210.

[58] A. Khandakar, M. E. H. Chowdhury, M. B. Ibne Reaz, S. H. M. Ali, M. A. Hasan, S. Kiranyaz, T. Rahman, R. Alfkey, A. A. A. Bakar, and R. A. Malik, "A machine learning model for early detection of diabetic foot using thermogram images," *Comput. Biol. Med.*, vol. 137, Oct. 2021, Art. no. 104838.

[59] A. Qayyum, A. Benzinou, M. Mazher, and F. Meriaudeau, "Efficient multi-model vision transformer based on feature fusion for classification of DFUC2021 challenge," in *Diabetic Foot Ulcers Grand Challenge*, M. H. Yap, B. Cassidy, and C. Kendrick, Eds. Cham, Switzerland: Springer, 2022, pp. 62–75.

[60] S. Ahmed and H. Naveed, "Bias adjustable activation network for imbalanced data," in *Diabetic Foot Ulcers Grand Challenge*, M. H. Yap, B. Cassidy, and C. Kendrick, Eds. Cham, Switzerland: Springer, 2022, pp. 50–61.

[61] 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, Feb. 2020, Art. no. 103616.

[62] 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 Informat. (BHI)*, Jul. 2021, pp. 1–4.

[63] M. Tan and Q. V. Le, "EfficientNetv2: Smaller models and faster training," in *Proc. Int. Conf. Mach. Learn.*, M. Meila and T. Zhang, Eds. 2021, pp. 10096–10106.

[64] H. Bao, L. Dong, S. Piao, and F. Wei, "BEiT: BERT pre-training of image transformers," 2021, *arXiv:2106.08254*.

[65] A.-K. Duong, H.-L. Nguyen, and T.-T. Truong, "Large margin cotangent loss for deep similarity learning," in *Proc. Int. Conf. Adv. Comput. Analytics (ACOMPA)*, Nov. 2022, pp. 40–47.

[66] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Classification of heart disease using K-nearest neighbor and genetic algorithm," *Proc. Technol.*, vol. 10, pp. 85–94, Jan. 2013.

[67] D. Sculley, G. Holt, D. Golovin, E. Davydov, T. Phillips, D. Ebner, V. Chaudhary, M. Young, J.-F. Crespo, and D. Dennison, "Hidden technical debt in machine learning systems," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 2503–2511.

[68] H. Kaur, H. S. Pannu, and A. K. Malhi, "A systematic review on imbalanced data challenges in machine learning: Applications and solutions," *ACM Comput. Surv.*, vol. 52, no. 4, pp. 1–36, Jul. 2020.

[69] D. Shanmugam, D. Blalock, G. Balakrishnan, and J. Guttag, "Better aggregation in test-time augmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 1214–1223.

[70] Z. Peng, L. Dong, H. Bao, Q. Ye, and F. Wei, "BEiT v2: Masked image modeling with vector-quantized visual tokenizers," 2022, *arXiv:2208.06366*.

[71] P. Pathak, J. Zhang, and D. Samaras, "Local learning on transformers via feature reconstruction," 2022, *arXiv:2212.14215*.

[72] L. Kwak and H. Bai, "The role of federated learning models in medical imaging," *Radiol., Artif. Intell.*, vol. 5, no. 3, May 2023, Art. no. e230136.

[73] N. Rieke, J. Hancox, W. Li, F. Milletarì, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, K. Maier-Hein, S. Ourselin, M. Sheller, R. M. Summers, A. Trask, D. Xu, M. Baust, and M. J. Cardoso, "The future of digital health with federated learning," *npj Digit. Med.*, vol. 3, no. 1, p. 119, Sep. 2020.

[74] M. F. Sohan and A. Basalamah, "A systematic review on federated learning in medical image analysis," *IEEE Access*, vol. 11, pp. 28628–28644, 2023.

[75] A. Rahman, M. S. Hossain, G. Muhammad, D. Kundu, T. Debnath, M. Rahman, M. S. I. Khan, P. Tiwari, and S. S. Band, "Federated learning-based AI approaches in smart healthcare: Concepts, taxonomies, challenges and open issues," *Cluster Comput.*, vol. 26, no. 4, pp. 2271–2311, Aug. 2023.

[76] G. A. Kaissis, M. R. Makowski, D. Rückert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," *Nature Mach. Intell.*, vol. 2, no. 6, pp. 305–311, Jun. 2020.

[77] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, M. Milchenko, W. Xu, D. Marcus, R. R. Colen, and S. Bakas, "Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data," *Sci. Rep.*, vol. 10, no. 1, p. 12598, Jul. 2020.

[78] M. Joshi, A. Pal, and M. Sankarasubbu, "Federated learning for healthcare domain-pipeline, applications and challenges," *ACM Trans. Comput. Healthcare*, vol. 3, no. 4, pp. 1–36, Oct. 2022.

[79] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2014, pp. 1–9.

[80] S. Kim and S. Lee, "Self-supervised augmentation of quality data based on classification-reinforced GAN," in *Proc. 17th Int. Conf. Ubiquitous Inf. Manag. Commun. (IMCOM)*, Jan. 2023, pp. 1–7.

[81] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, "Explainability for artificial intelligence in healthcare: A multidisciplinary perspective," *BMC Med. Informat. Decis. Making*, vol. 20, no. 1, pp. 1–9, Dec. 2020.

[82] L. Rubinger, A. Gazendam, S. Ekhtiari, and M. Bhandari, "Machine learning and artificial intelligence in research and healthcare," *Injury*, vol. 54, pp. 69–73, May 2023.

[83] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, and R. Ranjan, "Explainable AI (XAI): Core ideas, techniques, and solutions," *ACM Comput. Surveys*, vol. 55, no. 9, pp. 1–33, Sep. 2023.

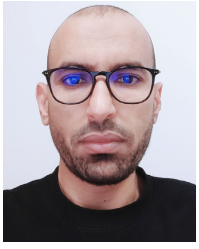
[84] D. Saraswat, P. Bhattacharya, A. Verma, V. K. Prasad, S. Tanwar, G. Sharma, P. N. Bokoro, and R. Sharma, "Explainable AI for healthcare 5.0: Opportunities and challenges," *IEEE Access*, vol. 10, pp. 84486–84517, 2022.

[85] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?' Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 1135–1144.

[86] E. T. Joa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4793–4813, Nov. 2021.

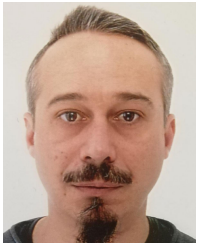
MOHAMMUD SHAAD ALLY TOOFANEE is currently pursuing the Ph.D. degree in artificial intelligence with the Research Laboratory XLIM, UMR CNRS 7252, University of Limoges, France. He is also a Senior Lecturer with Université des Mascareignes, Mauritius (UDM). He is the co-director of the master's course in artificial intelligence and robotics. His research interests are using machine learning in the field of health and more precisely diabetes prevention, management, and education. He is investigating the use of vision transformer for image and transformers for natural language processing.

SABEENA DOWLUT received the Ph.D. degree in health literacy from Université de la Réunion. She is a Senior Lecturer with Université des Mascareignes in Mauritius (UDM) and the Head of the Department of Applied Computer Science. She is also the co-director of the master's course in health and AI which will be starting in September 2023. She is co-supervising a Ph.D. thesis in the field of the IoT and health. She is also a member of "RÉseau Francophone de LIttératie en Santé (RÉFLIS)" and the Scientific Committee of AUF for the Indian Ocean and African Austral region.



MOHAMED HAMROUN received the Engineering degree from the Higher Institute of Computer Science and Multimedia, University of Sfax, Tunisia, in 2014, and the Ph.D. degree in computer science from the LaBRI Laboratory, University of Bordeaux, in December 2019. He pursued a career in industry before returning to school. He has held three postdoctoral research positions in two different laboratories, such as first with the LISIC Laboratory, University of Littoral; and second with the XLIM Laboratory, University of Limoges. He is an Associate Professor with the 3IL School and a Researcher with the XLIM Laboratory, UMR CNRS 7252, University of Limoges. His research interests are related to image and AI for health, pattern recognition, image and video processing, indexing and information retrieval, information extraction, machine learning, clustering, and deep learning.

KARIM TAMINE is an Associate Professor/a Researcher with the XLIM Laboratory, University of Limoges. He is the main resource person in artificial intelligence for the setting up of the master's degree with Université des Mascareignes Mauritius. He has supervised eight Ph.D. thesis and is supervising two Ph.D. students. His research work focuses on the use of artificial intelligence methods in various fields, such as computer graphics, security, and quality of service in dynamic communication networks.



VINCENT PETIT is currently pursuing the M.Sc. degree with ESPOIR, Université des Mascareignes, Mauritius. He is also a Software Engineer with Kaizen Solutions, Grenoble, France. He studied data science after a diverse experience in applied mechanics. His research interests include statistics, uncertainty assessment, machine learning, and deep learning, more especially computer vision and NLP, based upon which he strives to run research projects in the field of public health.

ANH KIET DUONG received the bachelor's degree (Hons.) from the Ho Chi Minh City University of Science, Vietnam. He is currently pursuing the master's degree with CRYPTIS, University of Limoges. His research interests include similarity learning, artificial intelligence, network security, federated learning, and cryptography.

DAMIEN SAUVERON is a Professor/a Researcher with the XLIM Laboratory, University of Limoges. He is also the Dean of the Faculty of Science and Technology. His research interests are related to smart card applications and security (at hardware and software level), RFID/NFC applications and security, mobile networks (e.g., UAV fleets) applications and security, sensors network applications and security, smart home applications and security, Internet of Things (IoT) security, cyber-physical systems security, security of distributed objects and systems, and security evaluation/certification processes.