

## Article

# A Deep Learning Method for Early Detection of Diabetic Foot Using Decision Fusion and Thermal Images

Khairul Munadi <sup>1,2</sup>, Khairun Saddami <sup>1</sup>, Maulisa Oktiana <sup>1</sup>, Roslidar Roslidar <sup>1,2</sup>, Kahlil Muchtar <sup>1,2</sup>, Melinda Melinda <sup>1</sup>, Rusdha Muharar <sup>1</sup>, Maimun Syukri <sup>3</sup>, Taufik Fuadi Abidin <sup>4</sup> and Fitri Arnia <sup>1,2,\*</sup>

<sup>1</sup> Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; khairul.munadi@unsyiah.ac.id (K.M.); khairun.saddami@unsyiah.ac.id (K.S.); maulisa.oktiana@unsyiah.ac.id (M.O.); roslidar@unsyiah.ac.id (R.R.); kahlil@unsyiah.ac.id (K.M.); melinda@unsyiah.ac.id (M.M.); r.muharar@unsyiah.ac.id (R.M.)

<sup>2</sup> Telematics Research Center, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia

<sup>3</sup> Medical Faculty, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; maimun@unsyiah.ac.id

<sup>4</sup> Department of Informatics, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; taufik.abidin@unsyiah.ac.id

\* Correspondence: f.arnia@unsyiah.ac.id

**Abstract:** Diabetes mellitus (DM) is one of the major diseases that cause death worldwide and lead to complications of diabetic foot ulcers (DFU). Improper and late handling of a diabetic foot patient can result in an amputation of the patient's foot. Early detection of DFU symptoms can be observed using thermal imaging with a computer-assisted classifier. Previous study of DFU detection using thermal image only achieved 97% of accuracy, and it has to be improved. This article proposes a novel framework for DFU classification based on thermal imaging using deep neural networks and decision fusion. Here, decision fusion combines the classification result from a parallel classifier. We used the convolutional neural network (CNN) model of ShuffleNet and MobileNetV2 as the baseline classifier. In developing the classifier model, firstly, the MobileNetV2 and ShuffleNet were trained using plantar thermogram datasets. Then, the classification results of those two models were fused using a novel decision fusion method to increase the accuracy rate. The proposed framework achieved 100% accuracy in classifying the DFU thermal images in binary classes of positive and negative cases. The accuracy of the proposed Decision Fusion (DF) was increased by about 3.4% from baseline ShuffleNet and MobileNetV2. Overall, the proposed framework outperformed in classifying the images compared with the state-of-the-art deep learning and the traditional machine-learning-based classifier.



**Citation:** Munadi, K.; Saddami, K.; Oktiana, M.; Roslidar, R.; Muchtar, K.; Melinda, M.; Muharar, R.; Syukri, M.; Abidin, T.F.; Arnia, F. A Deep Learning Method for Early Detection of Diabetic Foot Using Decision Fusion and Thermal Images. *Appl. Sci.* **2022**, *12*, 7524. <https://doi.org/10.3390/app12157524>

Academic Editor: Jan Egger

Received: 17 June 2022

Accepted: 19 July 2022

Published: 26 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Diabetes mellitus (DM) is one of the major diseases leading to death worldwide. Diabetes is a combination of metabolic illnesses marked by hyperglycemia—a condition of high sugar or glucose concentration in blood- resulting from abnormal insulin secretion, insulin action, or both. The chronic hyperglycemia of diabetes is linked to long-term damage, dysfunction, and failure of various organs, particularly the eyes, kidneys, nerves, heart, and blood vessels. Diabetic foot ulcer is one of the major complications caused by diabetes mellitus (DM) [1]. Foot ulcers occur due to damaged skin tissues under the big toes and on the foot's plantar. Foot ulcers cause the underlying skin layer to be exposed, affecting the feet to the bone as a result of the complication of uncontrolled diabetes. Improper and late handling of patients with a diabetic foot can result in an amputation of the foot.

All diabetic patients can suffer from diabetic foot ulcers; however, the ulcers are preventable with good foot care prior to chronic complications. Thus, an early detection of

diabetic foot ulcer should be provided. To date, various early detection methods have been proposed. Some worth mentioning are the works conducted by Saminathan et al. [2] and Usharani et al. [3]. However, the proposed methods are less effective, with a maximum accuracy of 96%.

In addition, various intelligent telemedicine monitoring systems have been employed to detect diabetic foot ulcer complications automatically [4,5]. Several approaches have been deployed using the medical image modality of visual images and thermal images. Thermal imaging is one of the most useful approaches in this case. It utilizes foot temperature to analyze foot complications. It is also a non-invasive method and does not cause harm to the body tissue [6].

Meanwhile, some studies have reported diabetic foot ulcer (DFU) classification using conventional machine learning [2,7,8]. Support vector machine (SVM) is one type of conventional machine learning techniques used for this purpose. However, according to the work by Vardasca et al. [9], SVM only achieved 87.5% accuracy using a private dataset. The SVM obtained a recall of 50%, which means only a half-positive patient can be classified as positive. Artificial Neural Networks (ANN) and k-nearest neighbors (kNN) were also applied to detect diabetic foot, yet they obtained lower results than those of SVM. Because the conventional method is still unable to detect and classify diabetic foot effectively, experts developed deep learning as a more effective method used for an image classification task [9].

Recently, experts have developed a classifier by implementing deep learning, as it shows capability in learning the image feature automatically. Several deep learning approaches that had been proposed in DFU detection are DFUNet [8], DFUQUTNet [10], ComparisonNet [11], and Segmentation [12,13]. However, most of the methods were proposed for the visual (visible) images, except the work by Cruz-Vega et al. [7], which used thermal images. They [7] proposed a new deep learning architecture to classify five classes of diabetic foot ulcers in DM groups only. It focused only on the risk levels of DFU patients, not on early detection. A previous study on early detection for DFU using thermal images was published by [14]. However, it was reported that an accuracy of only 97% was achieved using a combination of adaboost and random forest (RF) methods.

In the current study, we attempt to develop a new deep learning framework for the early detection of diabetic foot using decision fusion and thermal images. Generally, it is difficult to create a single classifier that works effectively on all of the testing data [15]. Therefore, decision fusion can be one of the solutions. Decision fusion can be applied to enhance the generalization of a training model and to avoid any biases in the classification results, particularly when it is trained using a limited dataset.

Pre-trained models of Heavyweight-based models, including ResNet101, DenseNet201, and XceptionNet, and Lightweight-based models, such as MobileNetV2, ShuffleNet, and EfficientNetB0, were trained on thermal plantar images. MobileNetV2 and ShuffleNet demonstrated good performance in balancing the latency and accuracy of the classification task; thus, these models were chosen in this work. Recently, most CNN architectures were proposed to increase the classification accuracy by adding more layers. As a result, the latency increased. Thus, it requires a large resource to obtain a CNN model. Meanwhile, the linear bottleneck proposed in MobileNetV2 prevents much information loss due to the non-linear curve. On the other hand, ShuffleNet, which proposes group channel shuffling, allows the information of each channel to flow and be fully connected. It also strengthens the representation of features. Hence, those two types of superiority are connected and complemented with each other by decision fusion to achieve better classification results.

The main contributions of this article are:

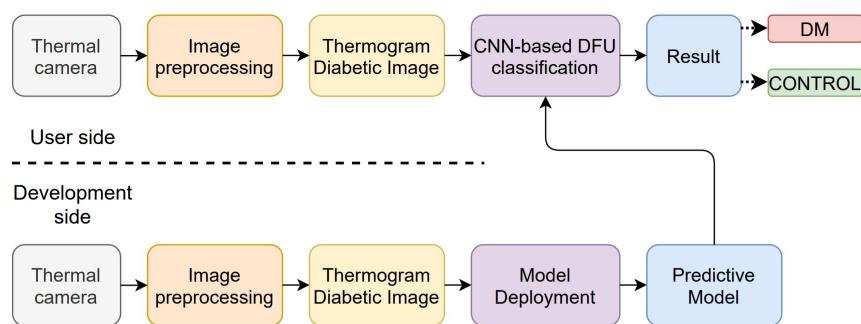
- Proposing a framework for the early detection of diabetic foot based on thermal images;
- Proposing a new decision rule for diabetic foot classification based on thermal images;
- Investigating the best fusion model for a high-accuracy DFU classification system based on pre-trained CNN models.

The rest of this article is organized as follows. Section two presents a computer-assisted system for a thermal-based diabetic foot classification. Section three provides the literature review. Section four explains the proposed framework while section five presents the simulation results and discussion. Finally, section six concludes the work.

## 2. Computer-Assisted System for Thermal-Based Diabetic Foot Classification

Humans are prone to making mistakes when diagnosing medical photographs due to several factors such as being over-worked [16]. In addition, limited human visual perception and optical illusions may impact the accuracy of a diagnosis [17]. Therefore, computer-assisted diagnostic (CAD) technology is created as a diagnostic-assisted system for clinicians. Furthermore, CAD technology can help healthcare organizations with only a few clinicians.

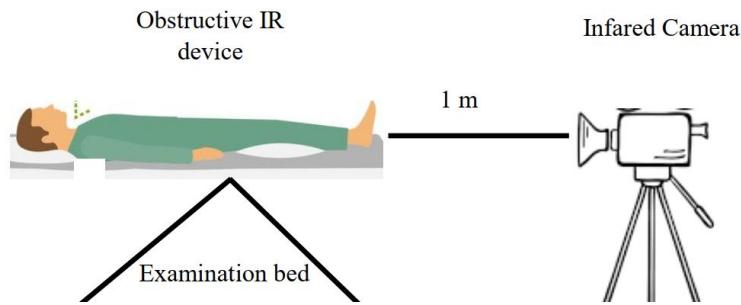
Recently, several studies proposed an image processing approach to enhance diagnostic accuracy [18–20]. For diabetic foot ulcer classification, a schema of the CAD system was developed as shown in Figure 1. The system has two sides: development and the user.



**Figure 1.** Computer-Assisted Diagnostic (CAD) system for diabetic foot ulcer (DFU) classification.

On the development side, the images are collected from patients with various conditions. The images are collected based on an acquisition system proposed by Hernandez et al. [21], as shown in Figure 2. The collected images are then pre-processed with image processing techniques such as denoising or segmenting the foot area. After pre-processing, the images are saved as a thermogram dataset for training purposes. The next step is model deployment for diabetic foot ulcer classification. Machine learning, including conventional machine learning and deep learning, is then trained on the DFU thermogram dataset. When performing conventional machine learning, the developer should train the model in two additional stages, namely, feature engineering and algorithm selection. Choosing the appropriate features for a specific algorithm is a challenging task in the classification process. For a deep learning approach, including convolutional neural networks (CNN), architectures and hyperparameter selection are the most determining processes in classifying the DFU images. The result of the model deployment stage is a predictive model that will be used on the user's side.

On the user's side, the user utilizes a thermal camera to acquire the diabetic foot. Then, the image is pre-processed to enhance the information in the image and remove any unnecessary data. Next, the image is classified by using the trained predictive model, and the model will label the result as either DM or CONTROL.



**Figure 2.** The image acquisition system based on work by Hernandez-Contreras et al. [21].

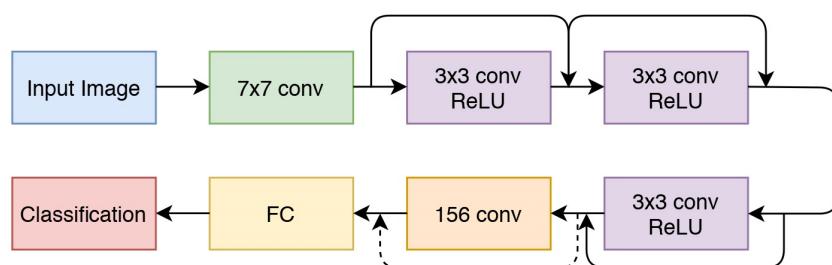
### 3. Literature Review

In this section, we present the benchmark of convolutional neural network (CNN) models and the decision fusion strategy. CNN has been widely researched since AlexNet showed a remarkable performance in the 2012 ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) competition. Since then, CNN has undergone rapid development. ResNet, the most popular CNN model, shows a very significant improvement in the classification performance [22]. Most ILSVCR winners were inspired by ResNet. The decision fusion strategy generalizes the models by combining or fusing the classification results to improve the classification performance.

#### 3.1. Convolutional Neural Network Models

##### 3.1.1. ResNet

ResNet is one of the deep learning architectures developed by utilizing the residuals of the previous layers. The core idea of ResNet is to keep the identity of the previous layer so that the mapping of each weighted layer will stay identical. ResNet stacks residual layers by performing an element-wise addition after every two convolutional layers [22]. Figure 3 shows an example of ResNet architecture. ResNet achieves a highly superior performance compared with the previous deeper architectures. Previous studies showed that deeper architectures result in a saturated accuracy during the training progress [22]. The residual layer proposed by ResNet is able to solve this problem.



**Figure 3.** The ResNet Architecture.

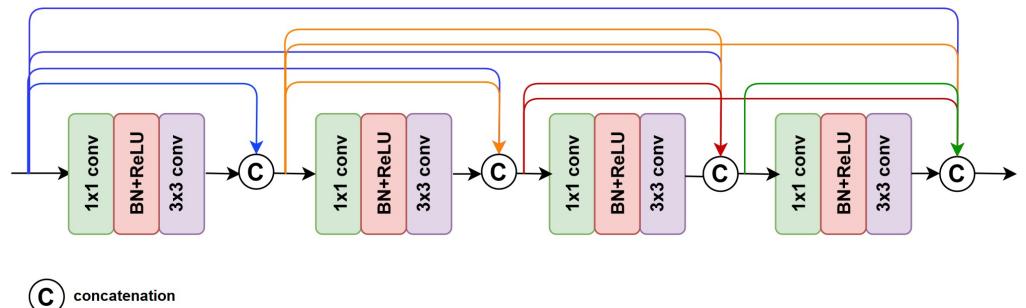
##### 3.1.2. DenseNet

DenseNet is a very deep and dense convolutional neural network based on connected residual layers [23]. DenseNet was acclaimed for its ability to improve gradient flow information using direct connections of entire sub-block layers. Unlike ResNet, which receives information only from the previous layers, DenseNet receives the information flow from all sub-blocks or sub-sequence layers, as represented in Equation (1) [23]:

$$X_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (1)$$

where  $[x_0, x_1, \dots, x_{l-1}]$  refers to the concatenation of the feature-maps produced in layers  $0, \dots, l-1$ . DenseNet arranges a sub-block by stacking a Batch Normalization, a ReLU,

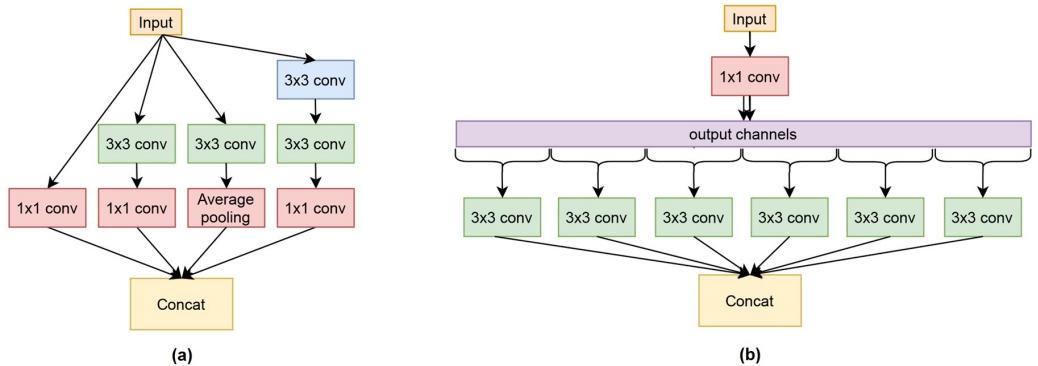
and a  $3 \times 3$  convolution followed by a transition block. The transition blocks are composed of a BN layer,  $1 \times 1$  convolutional layer, and a  $2 \times 2$  average pooling layer. DenseNet201 is the deepest DenseNet version proposed in the original paper. Figure 4 shows the Dense module that constructed the DenseNet.



**Figure 4.** The basic architecture of DenseNet.

### 3.1.3. XceptionNet

Xception [24] is an extension of an inception CNN architecture-based model that was introduced by GoogleNet [25]. Xception has more inception modules than the inception baseline, as shown in Figure 5. Figure 5a illustrates the inception-based line module, and Figure 5b shows the Xception module that constructed XceptionNet. Even though it has more modules, each of the Xception modules is more simple, and it also uses depthwise separable convolution to perform feature mapping.



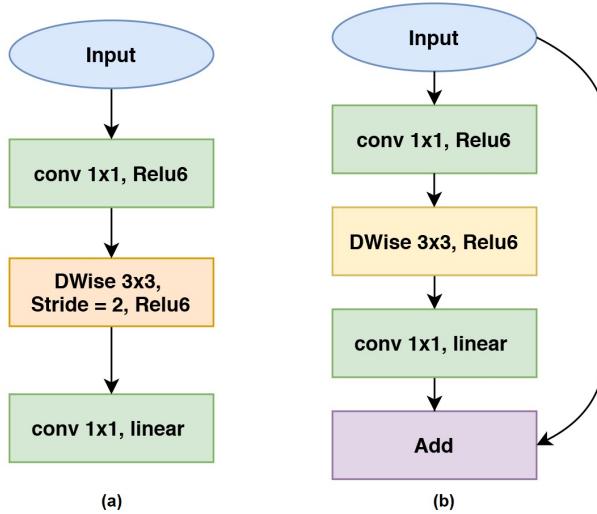
**Figure 5.** The basic architecture of XceptionNet: (a) Inception module, (b) XceptionNet module.

### 3.1.4. MobileNetV2

MobileNetV2 is a lightweight deep neural network designed for mobile-based applications [26]. MobileNetV2 proposes a residual inverted layer to improve gradient propagation in backpropagation progress and linear bottlenecks to reduce convolutional dimensionality. MobileNetV2 also uses ReLU6, as introduced in MobileNetV1. ReLU6 is a clipped ReLU in which the highest value is restricted to 6 while values lower than 0 will be kept as 0. ReLU6 is formulated in Equation (2):

$$y = \min(\max(x, 0), 6) \quad (2)$$

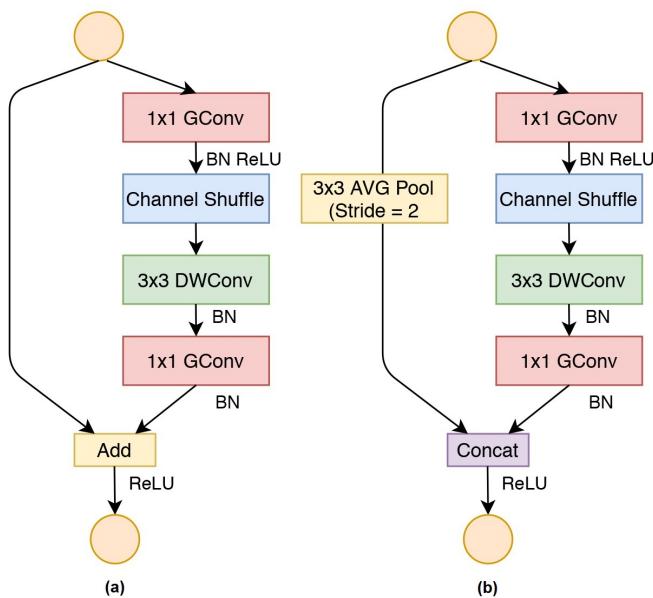
while  $y$  and  $x$  are the activated and pre-activated features, respectively. Figure 6 shows the basic architecture of MobileNetV2. MobileNetV2 has two main structure blocks. The structure shown in Figure 6a is the main structure of MobileNetV2, while the structure shown in Figure 6b is used when the dimension reduction stage is being applied.



**Figure 6.** The basic architecture of MobileNetV2: (a) MobileNetV2 units with stride = 1, (b) MobileNetV2 units with stride = 2.

### 3.1.5. ShuffleNet

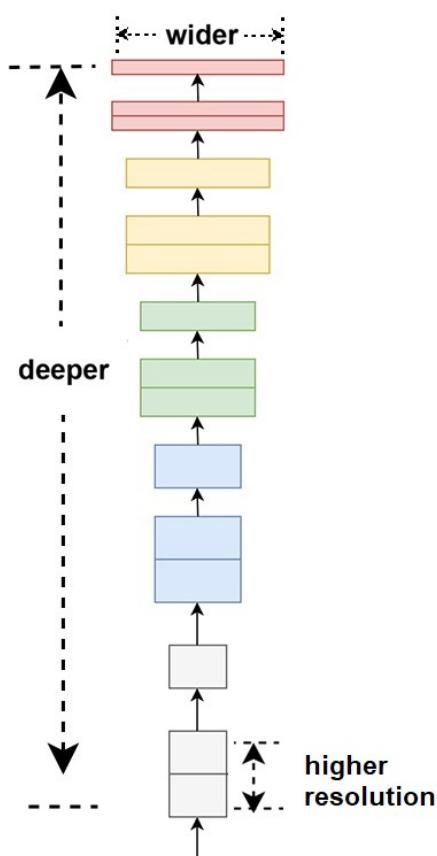
ShuffleNet proposes channel shuffling and group convolution layers to reduce the computational cost and complexity [27]. The purpose of channel shuffling is to shuffle and cross the features from channels so that the feature weights will be the combination of the channels. In the proposed framework, we shuffle every four-channel features in the channel-shuffling layers. This operation leads to a superior structure in the CNN architecture due to multiple feature weights existing in the convolutional layers [27]. Furthermore, the channel shuffling layers are combined with the group convolution layers to achieve a more powerful CNN model. Group convolution layers were introduced in AlexNet [28] and used with different approaches by [29,30] and certainly in MobileNetV1 [31], which is the fundamental architecture of ShuffleNet. Figure 7 shows the basic architecture of ShuffleNet. Figure 7a is the ShuffleNet unit with stride 1 for dimension reduction, while Figure 7b is the ShuffleNet unit with stride 2.



**Figure 7.** The basic architecture of ShuffleNet: (a) ShuffleNet units with stride = 1, (b) ShuffleNet units with stride = 2.

### 3.1.6. EfficientNet

EfficientNet proposes a new concept of scaling convolutional layers to extract deeper features called the compound scaling method [32]. While other CNN approaches focus on finding the best architecture layers, EfficientNet's approach is to investigate the layer's depth and tensor size. Hence, EfficientNet can also be implemented in other CNN architectures' base-lines, such as ResNet and MobileNet. Figure 8 shows a compound scale model concept used in EfficientNet. EfficientNet has eight family architectures which are deployed in compound scaling model. The basic composer of EfficientNet is known as EfficientNetB0. It is further developed, and its latest variant is EfficientNetB7.



**Figure 8.** The basic architecture of EfficientNet.

### 3.2. Decision Fusion

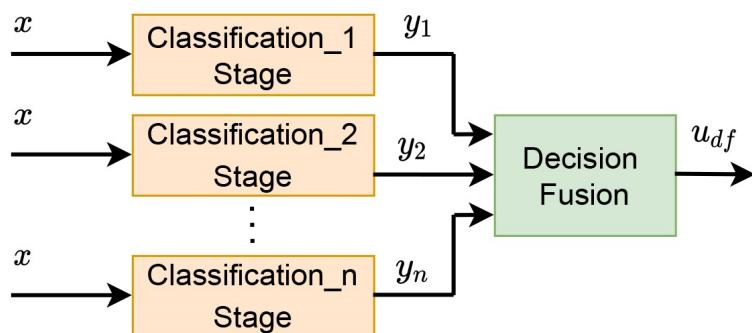
Fusing the data information is one way to achieve the best classification performance in a pattern recognition task. Based on the level of data processing, fusion strategies can be categorized into three main classes: low-level or data fusion, intermediate-level or feature fusion, and high-level or decision fusion.

In the first main class, or data fusion, data from different sources are combined to employ a new data source. Combining several data sources will add more valuable information. For example, fusing NIR infrared images and visible images will enrich the information in face recognition. In the second main class, or feature fusion, valuable features are selected and combined. In the third class, or decision fusion, several classification results from each classifier are combined to achieve the final result.

Decision fusion is defined as one of the fusion schemes where the decisions of multiple classifiers are combined into a final decision [33]. For simplification, the decision fusion is illustrated in Figure 9, where  $x$  represents input images classified using different classifica-

tion techniques.  $y_1$  and  $y_2$  are the image classification results using different classification methods. The final decision of this result is defined as  $u_{df}$ .

Decision fusion has been used to enhance and generalize the performance of a machine learning model. Oszust et al. [34] used a decision fusion technique to optimize the image quality assessment (IQA) methods. Fusing several IQA methods will improve the accuracy of the image quality assessment results. Furthermore, Zhang [35] proposed a decision fusion of Markov random field (MRF) and convolutional neural network (CNN) to enhance the resolution and quality of remote sensing images. In their work, MRF and CNN classified positive and negative results by handling them separately, in which MRF handled the negative region while CNN handled the positive one. The negative and the positive region are correct and incorrect classification results of the remote sensing image pixels, respectively. Score-level fusion is proposed to combine the classification results of Gaussian Mixture Model (GMM), support vector machine (SVM), and a combination of GMM-SVM [36].



**Figure 9.** A decision fusion scheme.

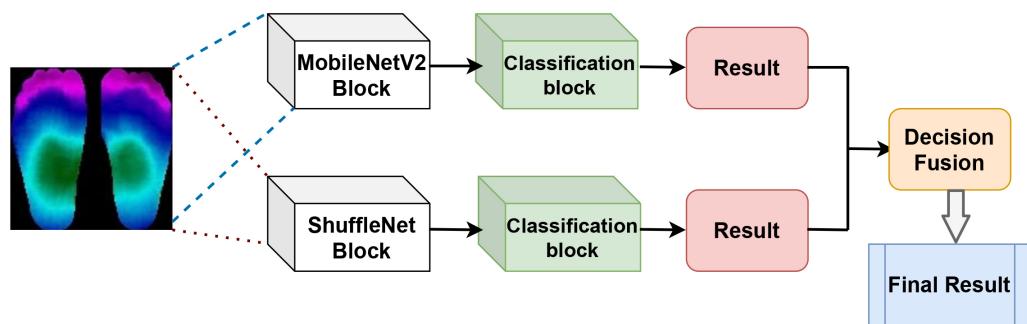
A large number of studies reported the decision fusion performance of CNN-based classification methods [37–41]. Abdi et al. [38] proposed the spectral-spatial decision fusion of CNN classification results to obtain the best classification result based on remote sensing image data. On the other hand, Rwigema et al. [40] introduced decision fusion for age and gender classification based on conventional neural network and CNN. In the paper, it was confirmed that CNN has superior performance compared with that of conventional machine learning. The Behavior Knowledge Space (BKS) achieved a better fusion classification result compared with that of Naïve Bayes (NB) fusion, whereas the Sum rule decision fusion achieved a higher result compared with those of the majority of voting and Naïve Bayes.

Based on the finding by James et al. [37], medical image fusion can result in notable performance in medical image processing. Several papers reported that fusing the information of medical images can result in better performance compared with that of a single classification system [39]. In recent years, decision fusion has also been applied in deep learning for multi-purpose medical imaging [41]. Averaging, majority voting, and multi-FCN fusion based are the most widely used decision fusion methods in medical image processing. As proposed by Khandakar et al. [14], the combination of AdaBoost and Random Forest results in 97% accuracy in classifying DFU thermal images.

#### 4. The Proposed Framework

In this paper, we proposed a novel framework to classify diabetic foot ulcer (DFU) using a decision fusion rule and thermal images. The proposed method used the DFU images to train the pre-trained benchmarking CNN architectures separately, namely, heavyweight models, which are ResNet, DenseNet, XceptionNet, and lightweight models including ShuffleNet, MobileNetV2, and EfficientNet. Most of the pre-trained models have been proven to achieve superior performance in biomedical thermal image classification [42].

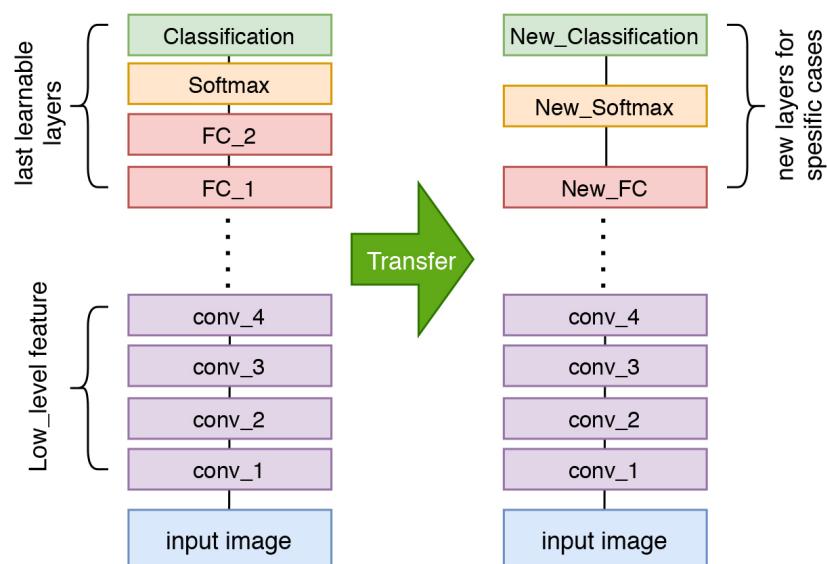
The pre-trained models had been trained on the ImageNet dataset and had very appropriate weights. Figure 10 shows our proposed framework based on MobileNetV2 and ShuffleNet. To enhance the performance of the two models, we fused the results using a decision fusion method. The implementation of the proposed method is presented in Algorithm 1 in Section 4.2.



**Figure 10.** The proposed framework.

#### 4.1. Transfer Learning

Transfer learning refers to a concept in deep learning where a model that has been trained using a particular dataset is used for the classification process of a new dataset with limited images. The model that has been trained and will be used for a new classification task is called a pre-trained model. The transfer learning concept is useful for solving issues related to insufficient datasets for the training process. Various models can be used as pre-trained models for transfer learning, i.e., ResNet, MobileNet, and ShuffleNet. These models have been trained using the ImageNet dataset, consisting of over 1 million images with 1000 classes, resulting in superior performance. The application of transfer learning or the pre-trained model in thermal image classification has shown satisfactory performance [42,43]. The simplest way to perform transfer learning is by replacing the layers after the last trainable layer with the proposed classification layer. Figure 11 illustrates the transfer learning process.



**Figure 11.** The transfer learning concept using a pre-trained model.

#### 4.2. Decision Rule

MobileNetV2 and ShuffleNet showed excellent performance in classifying medical images or thermal images [42,44]. In addition, both of the pre-trained models were developed based on the same fundamental architecture, namely, MobileNetV1. MobileNetV2 proposes inverted residuals and linear bottlenecks to improve MobileNetV1, while ShuffleNet proposes channel shuffling and group convolution to perform the task. ShuffleNet and MobileNetV2 were selected as the baseline classifiers based on the performance of the sensitivity and specificity rate of the mode itself prior implementing the fusion strategy. As neither of the models achieved a 100% accuracy rate, we performed the fusion strategy at the decision stage. In addition, ShuffleNet and MobileNetV2 have shown a reliable performance in classifying the medical images [20,42,44]. Other pre-trained benchmark models also showed very promising performances, but ShuffleNet and MobileNetV2 has less learning parameters and smaller model sizes.

In the present study, it was found that transferring the pre-trained model weight from ImageNet to the DFU dataset did not improve the classification performance significantly. Using the single-baseline CNN models did not achieve a 100% accuracy rate; therefore, to enhance the performance of CNN models in classifying DFUs, we fused the results of MobileNetV2 and ShuffleNet. ShuffleNet had the ability to classify the negative DFU image precisely (100% accuracy) but not for positive DFU images. On the other hand, MobileNetV2 had the ability to classify the positive DFU image precisely (100% accuracy) but not for negative DFU images. The proposed decision fusion is formulated as follows:

$$R(x) = \begin{cases} S(x), & \text{if } S(x) = \text{POSITIVE.} \\ M(x), & \text{otherwise.} \end{cases} \quad (3)$$

where  $R(x)$  is the decision fusion result of positive or negative patients;  $S(x)$  is the classification results using ShuffleNet;  $M(x)$  is the classification result using MobileNetV2. Decision fusion determines which result will be chosen as the output. Based on Equation (3), if a patient is found to have a positive result by ShuffleNet, the patient will be considered as a positive patient. In contrast, if a patient is found to have a positive result by MobileNetV2, the patient will be considered as a negative patient.

---

#### Algorithm 1: Decision Fusion Schema

---

```

input :A thermal DFU image
output:  $M(x)$  is the label result of the MobileNetV2 model;
           $S(x)$  is the label result of the ShuffleNet model;
           $R(x)$  is the label result of decision fusion;
The input image is classified using MobileNetV2 and ShuffleNet;
if  $S(x) == \text{POSITIVE}$  then
    |  $R(x) = S(x)$ ;
else
    |  $R(x) = M(x)$  ;
end

```

---

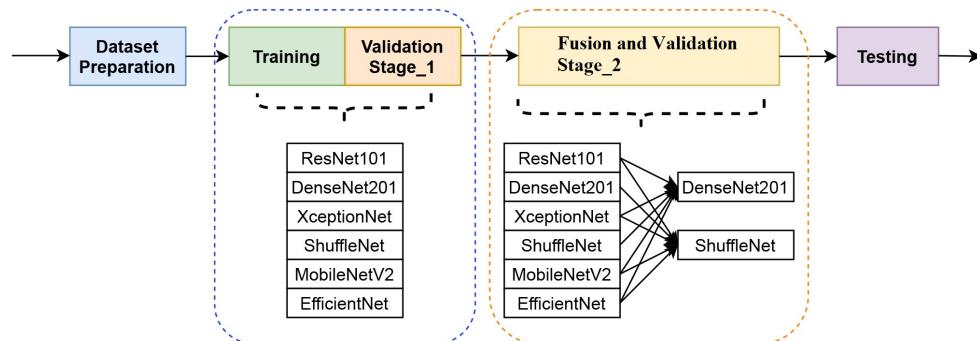
## 5. Simulation Result and Discussion

### 5.1. Simulation Setup

In this subsection, we will describe the simulation stages, namely: (1) dataset preparation, (2) training with validation stage 1, (3) fusion deployment and validation stage 2, and (4) testing, as shown in Figure 12.

In the first stage, the dataset was prepared. Then, in the second stage, each CNN model was trained using the transfer learning concept and validated. In the third stage, decision fusion was performed, and then the model was validated. Validation stage 2 aims to obtain the best CNN model fusion compared with the single-model performance. The

last stage is the testing stage to measure the model's performance. The model's performance is evaluated using the metric of accuracy, sensitivity, specificity, and F-measure. A detailed description is provided as follows.



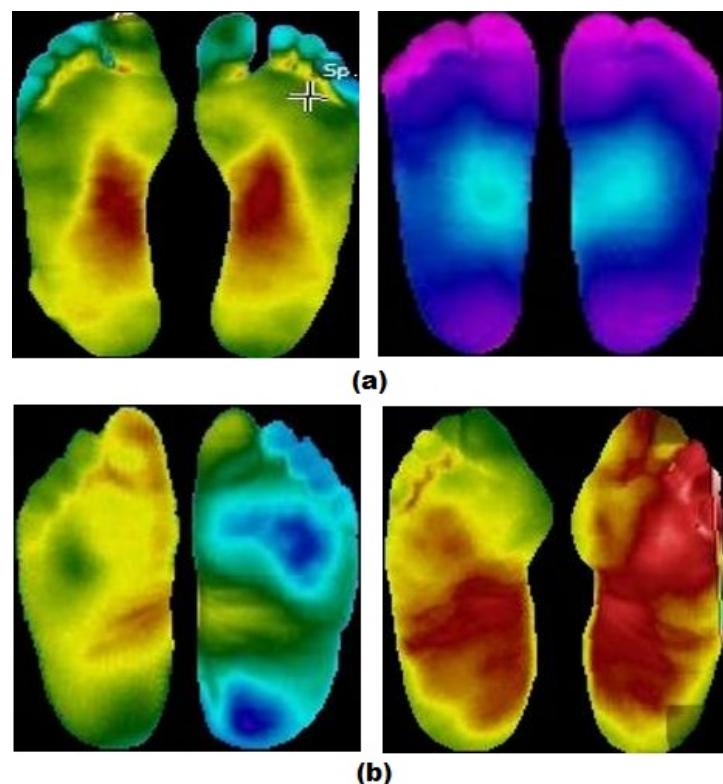
**Figure 12.** The simulation steps.

### 5.1.1. Dataset

We conducted the experiments on the plantar thermogram database provided by Hernandez-Contreas et al. [21]. The datasets were taken using a thermal camera and consisted of separate left and right foot images. As mentioned by Netten et al. [45], the diabetic foot depends on ipsilateral and contralateral temperature. To enable the CNNs to learn the disease, we put/combined the right foot image and the left foot image of each patient side by side, as shown in Figure 13. The images of positive diabetic patients were labeled as DM, while the negative ones were labeled as CONTROL. There were 142 positive DM patients and 45 negative CONTROL patients. Due to the limited number of datasets for deep learning and unbalanced classification problems emerging during the training process, we performed image augmentation by rotating, scaling, and translating the images, as performed in some of the literature [46,47]. Image augmentation resulted in 1200 images each of positive and negative DM patients. The augmentation was performed more on images of negative CONTROL patients to balance the dataset. To deal with the stages in the simulation setup, we split the dataset into 75:5:10:10, which means 75% for training, 5% for validation stage 1, 10% for validation stage 2, and 10% for testing. Because MobileNetV2 and ShuffleNet required  $224 \times 224$  pixels for an input image, we resized the dataset. Table 1 shows detailed numbers of images in the dataset before and after augmentation.

**Table 1.** Number of images before and after augmentation.

Class	Number of Images	
	Before Augmentation	After Augmentation
CONTROL	142	1200
DM	45	1200
<b>Total</b>	<b>187</b>	<b>2400</b>



**Figure 13.** Examples of DFU thermal data training: (a) the thermal images of a negative patient, (b) the thermal images of a positive patient.

#### 5.1.2. Training and Validation Stage One

We trained each pre-trained network on the plantar thermogram dataset in this step. Before training, we set up the learning parameters to reach maximum accuracy while maintaining the stability of the learning. We adopted the Stochastic Gradient Descent with Momentum (SGDM) optimizer with momentum in the range of 0.93 to 0.99. The learning rate was set between  $10^{-3}$  and  $10^{-4}$ . We trained the model using seven epoch values, namely, 10, 25, 50, 75, 100, 150, and 200. Each epoch has 500 batches.

At this stage, we validated the model performance as a single classification result for each model of the trained network. All model performances were analyzed, and the fusion step was determined based on this result.

#### 5.1.3. Fusion and Validation Stage Two

At this stage, the experiment aims to enhance the model's performance by fusing the classification results of every single model. The combination of models was determined based on the single classification result. The model with the highest accuracy rate was combined with other models to achieve the best performance. The model was then validated using a different validation dataset from the training dataset.

#### 5.1.4. Testing and Evaluation

We evaluated our proposed DF model and other baseline deep learning models. The testing dataset was fed into the developed model; then, we calculated the evaluation metrics. The first metric is the learning parameter, which refers to the number of weighted parameters learned and changed during the training process. The fewer the learning parameters, the faster the training process is. The second metric is accuracy, which is formulated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

in which  $TP$  is a true positive,  $TN$  is a true negative,  $FP$  is a false positive, and  $FN$  is a false negative.

The third metric is specificity. It represents the ability of a technique to classify a positive patient correctly. Specificity is calculated as follows:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (5)$$

The fourth one is recall or sensitivity. It represents the ability to classify a negative patient correctly. It is calculated as follows:

$$\text{Recall or sensitivity} = \frac{TP}{TP + FN} \quad (6)$$

Accuracy, specificity, and sensitivity are the main metrics used to evaluate deep learning performance. Other metrics are also used, namely, precision and F-Measure. Precision is computed using the following formula:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

Precision shows the level of a correctly classified positive patient from a totally classified positive. The last metric is F-measure (FM), which is the harmonic mean of recall and precision. FM is formulated as follows:

$$\text{F-measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \quad (8)$$

To assess the proposed method, we compared it with the origin of MobileNetV2, ShuffleNet, ResNet50, DFUNet [7], the combination of MobileNetV2 and ShuffleNet using the average pooling (MS+AP) [48], the combination of MobileNetV2 and ShuffleNet using max pooling (MS+MP), Local Binary Pattern (LBP) and support vector machine (SVM) classifier (LBP+SVM), Local Binary Pattern (LBP) and Naïve Bayes (NB) classifier (LBP+NB), and Local Binary Pattern (LBP) and random forest (RF) classifier (LBP+RF). We used 1000 trees for RF classification. In SVM classification, Sequential Minimal Optimization (SMO) [49] was applied with a linear kernel. Other kernels, namely, Gaussian and Polynomial, resulted in significantly poorer performance.

## 5.2. Training Results

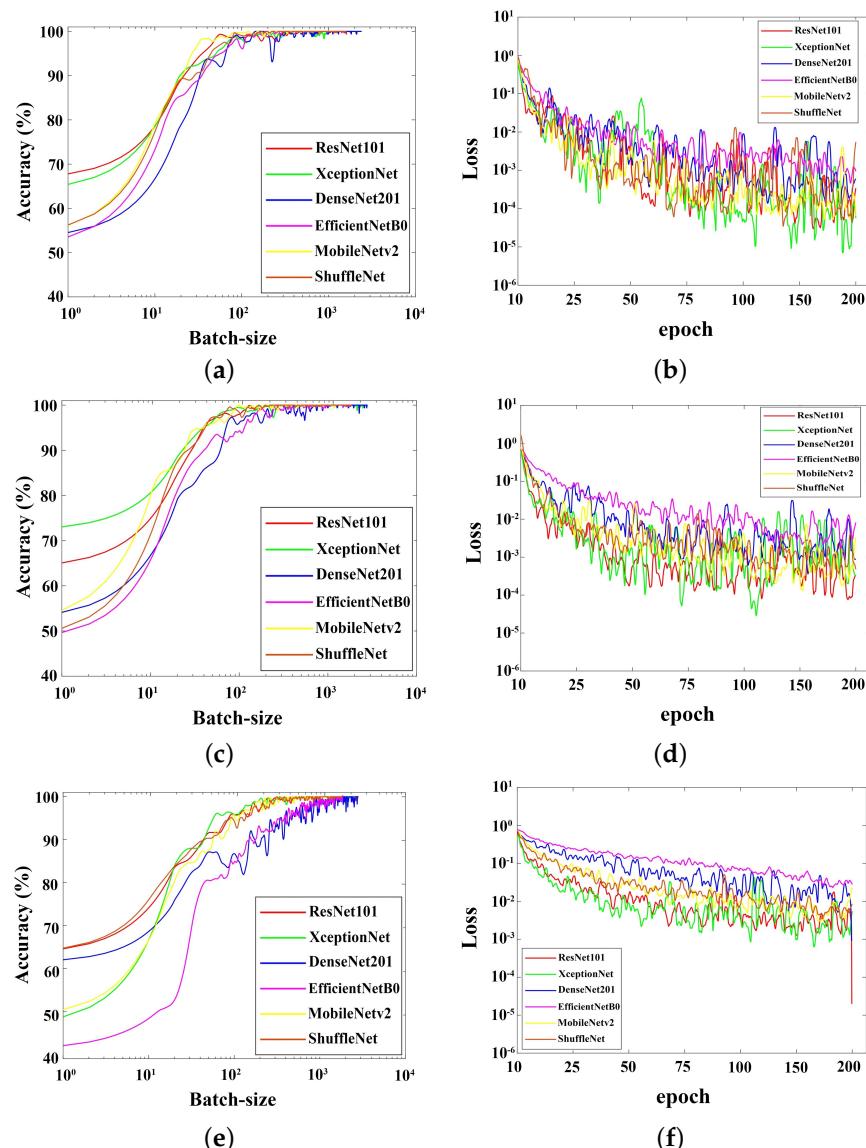
Figure 14 shows the accuracy and loss of training behaviors during the training process. Figure 14a shows the accuracy curves with the increments in batch-size. All networks were trained using the learning rate of  $10^{-3}$ . Among all, ResNet and XceptionNet tuned out to have the highest accuracy rates, while DenseNet had the lowest one. DenseNet was found to be the slowest CNN that reached convergence when trained on the thermal DFU images. Two mobile-based networks, MobileNetV2 and ShuffleNet, had similar accuracy values to the beginning of the training process.

In Figure 14c, the accuracy rates of ShuffleNet and DenseNet increased since the training started. DenseNet reached convergence more slowly than ShuffleNet and other networks. On the other hand, decreasing the learning rate decreased the starting accuracy of EfficientNet, MobileNetV2, and XceptionNet. Overall, DenseNet and EfficientNet showed slower convergence achievements. The graphs show that the higher the learning rate value. the faster the training accuracy towards convergence. Fluctuated training curves also inform that the dataset images has low-variance as a result of the augmentation process.

Figure 14b,d,f show the training loss values for six CNNs trained on the DFU dataset. Based on Figure 14b, the models trained using the learning rate value of  $10^{-3}$  showed high fluctuation trends. It possibly means that the variance within the dataset is low because the

dataset contains augmented data. XceptionNet showed unstable performance and hit a lower curve when the training epoch was between 25 and 50. However, the curve rose at from epochs 50 to 75 and dropped again at epochs above 75. The curve of ResNet seemed to follow that of the XceptionNet and experienced a similar shape change. The curves of EfficientNet and DenseNet looked more dominant at the top of the training loss curve. The MobileNetV2 and ShuffleNet curves are in the middle of the loss curve. Figure 14d shows the training loss curve at the learning rate of  $0.0005$  or  $5 \times 10^{-4}$ . The XceptionNet and ResNet curves hit the lowest points, just like the curves in Figure 14b.

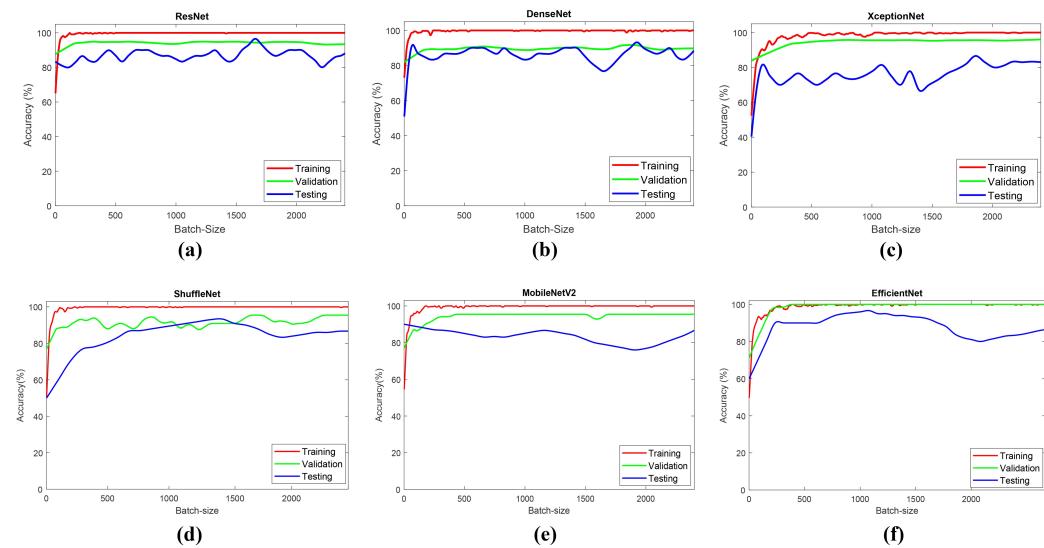
The curve shown in Figure 14f is a curve of the training models at the learning rate of  $10^{-4}$ . The EfficientNet model had the highest loss value during the training process, followed by that of DenseNet, MobileNetV2, ShuffleNet, ResNet, and XceptionNet. According to Figure 14b,d,f, the learning rates affected the movements of each model's weight towards convergence. The smaller the learning rates, the less the training loss values fluctuated. Oppositely, each curve in Figure 14a,c,e exhibited a more stable trend.



**Figure 14.** Training accuracy during the training process. (a,b): Trained using learning rate of  $10^{-3}$ . (c,d): Trained using learning rate of  $5 \times 10^{-4}$ . (e,f): Trained using learning rate of  $10^{-4}$ .

Figure 15 shows the learning curve of the CNN models. Based on the learning curve, all CNN models showed that the training accuracy increased to a stability point. The

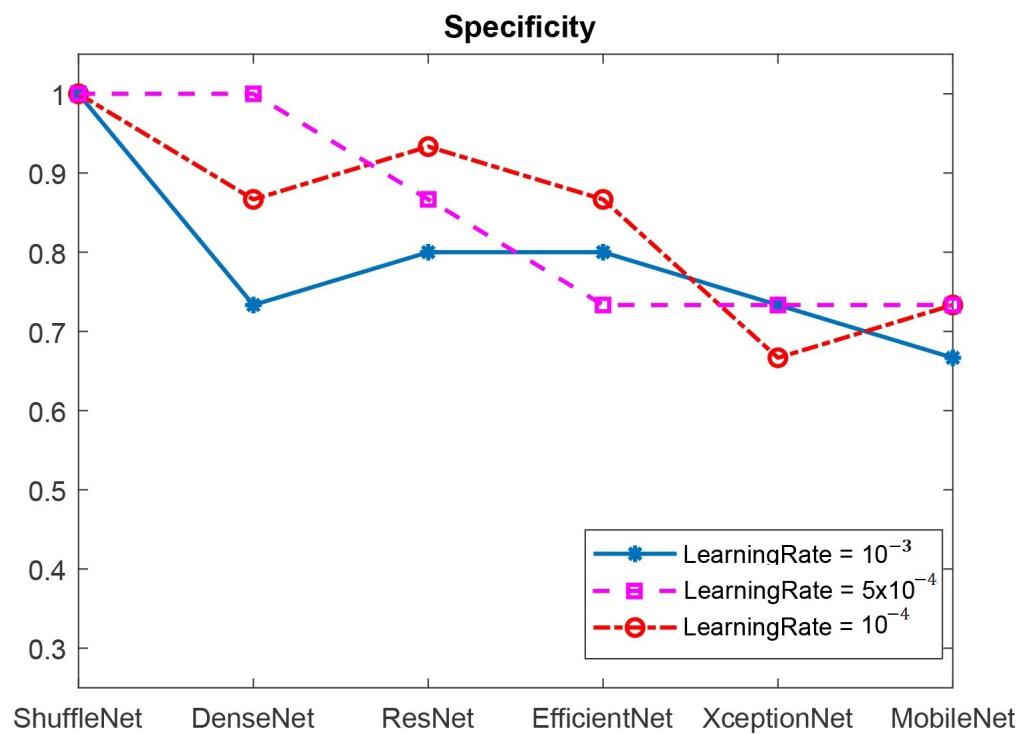
validation accuracy also showed similar trends. Every model in the simulation showed that the training and the validation accuracy has a small gap. Furthermore, the testing accuracy is under the training and validation accuracy. This trend indicated that the models have a good fit to the learning curve.



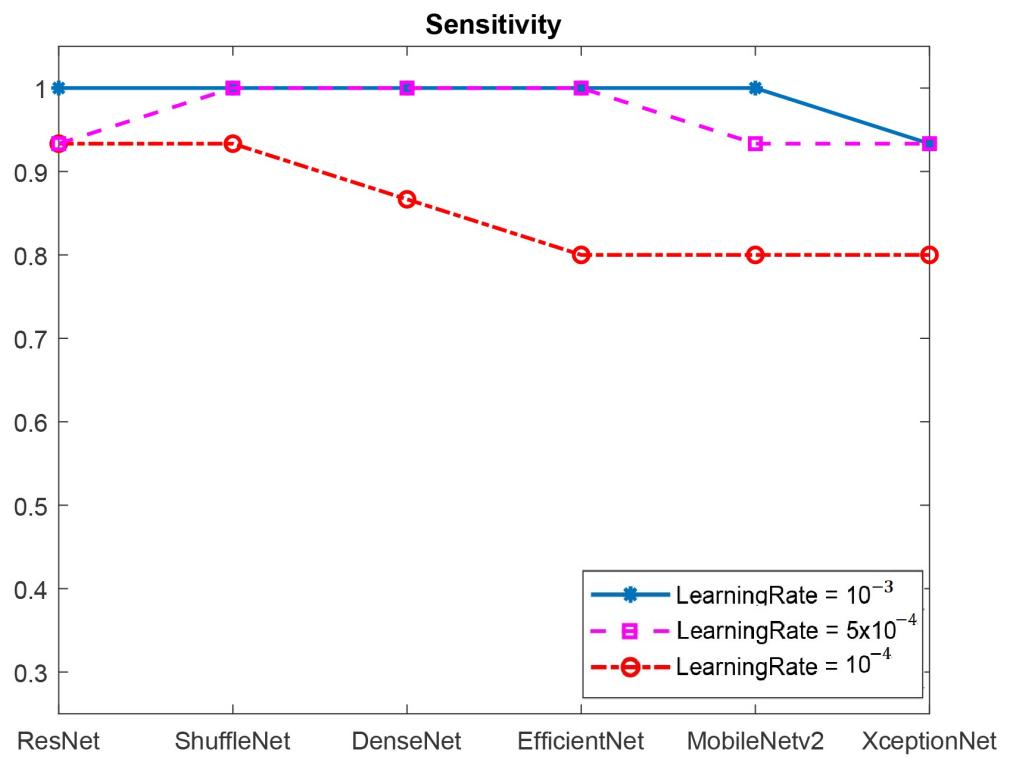
**Figure 15.** Learning curves of all simulated models: (a) ResNet; (b) DenseNet; (c) XceptionNet; (d) ShuffleNet; (e) MobileNetv2; (f) EfficientNet.

### 5.3. Single Classification Results

Figures 16 and 17 show the specificity and sensitivity of the CNN benchmark architectures, respectively. Based on Figure 16, ShuffleNet (with all learning rate values) and DenseNet (with a learning rate of  $5 \times 10^{-4}$ ) achieved a specificity of 1.0. ShuffleNet showed remarkable performance in classifying the positive images with specificity of 1.0 at all the learning rate values, whereas DenseNet reached specificity of 1.0 only at the learning rate of  $5 \times 10^{-4}$ . ResNet and EfficientNet showed lower specificity values compared to those of ShuffleNet and DenseNet. XceptionNet and MobileNet resulted in the poorest performance at all learning rates. The DFU image classification depends on thermal color distribution from the thermal images. As ShuffleNet works by shuffling the channel information of each colored layer, it keeps the information flow during the training process in the network. Other CNN models might lose information during the training and weighting process due to separable and unrelated channels from each layer.



**Figure 16.** The specificity values of six CNN benchmarking architectures.



**Figure 17.** The sensitivity values of six CNN benchmarking architectures.

#### 5.4. Decision Fusion Results

In Figure 17, it was shown that ResNet, Efficient-Net, ShuffleNet, and MobileNetV2 all had a sensitivity of 1.0 at the learning rate of  $10^{-3}$ . Therefore, we then fused ResNet,

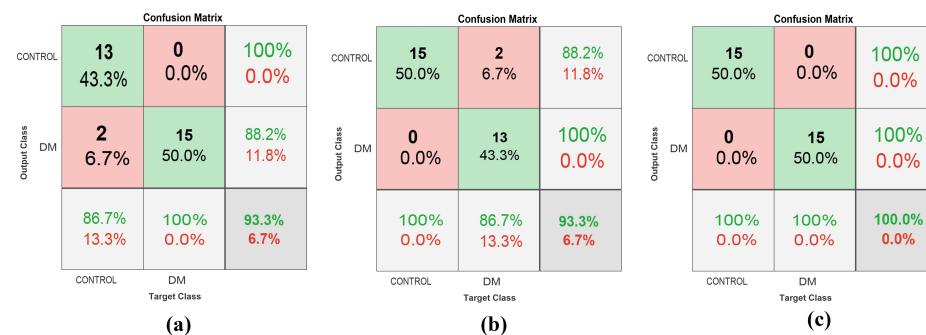
Efficient-Net, ShuffleNet, and MobileNetV2 with DenseNet and ShuffleNet, which had specificities of 1.0, to classify the DFU images. The results are shown in Table 2.

**Table 2.** Decision fusion performance for thermal DFU classification. Bold font indicates considerable results.

Deep Learning Based							
Fusion Methods	Learning Parameters	Accuracy	Sensitivity or Recall	Specificity	Precision	F-Measure	Model Size
ResNet + DenseNet	60.64 M	1.000	1.000	1.000	1.000	1.000	217.3 MB
Xception + DenseNet	38.94 M	0.967	0.934	1.000	1.000	0.967	140.4 MB
EfficientNet + DenseNet	22.11 M	1.000	1.000	1.000	1.000	1.000	80.8 MB
ShuffleNet + DenseNet	18.94 M	1.000	1.000	1.000	1.000	1.000	69.52 MB
MobileNetv2 + DenseNet	20.31 M	1.000	1.000	1.000	1.000	1.000	74.45 MB
ResNet + ShuffleNet	43.41 M	1.000	1.000	1.000	1.000	1.000	154.22 MB
DenseNet + ShuffleNet	18.94 M	1.000	1.000	1.000	1.000	1.000	69.52 MB
Xception + ShuffleNet	21.71 M	1.000	1.000	1.000	1.000	1.000	77.32 MB
EfficientNet + ShuffleNet	4.88 M	1.000	1.000	1.000	1.000	1.000	17.72 MB
<b>MobileNetv2 + ShuffleNet</b>	<b>3.08 M</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>11.37 MB</b>

Table 2 shows the decision fusion result of several models. Most of the models, including the proposed ones, achieved 100% accuracy rate. The proposed method is distinguished from other fusion models because the proposed model required far less learning parameters and has a smaller size. For example, the proposed model only has 3.08 M learning parameters and a 11.37 MB model size, while the fusion of MobileNetv2 and DenseNet reached 20.31 M learning parameters and a 74.45 MB model size (both are 6.5x larger than the proposed model).

According to the confusion matrix in Figure 18, it was confirmed that the proposed method demonstrated the best performance. When the model was only trained using a single model, the accuracy was only 93.34%. Nevertheless, when the model was fused, the accuracy increased to 100%. Based on the proposed method, the drawback of MobileNetv2 in classifying the CONTROL images was solved by ShuffleNet and vice-versa.

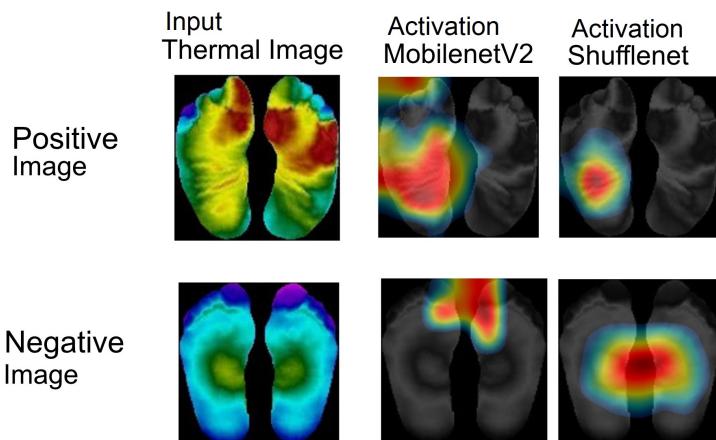


**Figure 18.** Confusion matrix of: (a) MobileNetv2; (b) ShuffleNet; (c) the proposed method.

Fusing MobileNetV2 and ShuffleNet resulted in a promising performance. This result is confirmed by visualizing the activation map generated by MobileNetV2 and ShuffleNet. Activation mapping layers were used to identify and illustrate a CNN model at a specific layer. Figure 19 shows the activation map of the last activation layer from MobileNetV2 and ShuffleNet. The upper row indicates positive images while the lower row indicates negative images.

According to Figure 19, positive and negative images generated by the last activation layer of MobileNetV2 and ShuffleNet had different mapping intensities. The intensity values of the positive patients' images were dominant in the contralateral region. The

images display an asymmetrical thermal distribution. On the other hand, thermal distributions of the negative patients' images are symmetrical. Thus, it can be concluded that MobileNetV2 and ShuffleNet can represent the thermal distributions of diabetic feet more effectively using the activation layer. Based on Figure 19, ShuffleNet provides a more focused representation than MobileNetV2 to indicate a positive or negative patient.



**Figure 19.** The activation function result of MobileNetV2 and ShuffleNet.

### 5.5. The Comparison of Several Methods' Performances

In Table 3, the performance of the proposed decision fusion model is compared with the performances of other CNN-based classification models and conventional machine learning models. The decision fusion between EfficientNet and ShuffleNet is indicated by ESNet. We compared the proposed model and the ESNet with Cruz-vega architecture [7], average pooling approach [48], ResNet101 and MobileNetV2, and ShuffleNet combination that was fused using Max-Pooling probability (MS+MP). The proposed model and the ESNet achieved a score of 1.0 for all evaluation metrics, i.e., accuracy, sensitivity or recall, precision, and F-Measure, which are the highest values compared with those of other methods. The proposed model and the ESNet have higher accuracy, sensitivity, precision, and F-measure compared with all CNN-based classification methods, except for Cruz-Vega, with a precision of 1.0. Furthermore, the proposed framework shows a better performance compared with those of the conventional machine-learning-based classification methods. The combination of local binary pattern (LBP) and a support vector machine (SVM) resulted in noticeable performance, so did the combination of LBP and random forest (RF). The LBP+SVM and LBP+RF methods obtained an accuracy of 0.880, a recall of 0.917, a precision of 0.917, and an F-Measure of 0.917. LBP combined with Naïve Bayes (NB) result in the poorest performance compared with those of deep learning and traditional machine-learning-based methods, with an accuracy of 0.710, a recall of 0.890, a precision of 0.670, and an FM of 0.760.

Table 3 also shows a comparison of the learning parameters used in the deep learning models, including in the proposed model. The proposed model is composed of MobileNetV2 and ShuffleNet, which means that it has the learning parameters of both architectures. The proposed model has 3.08 M learning parameters, while the ESNet has 4.88 M. The reason why the proposed model has fewer learning parameters is because it is composed of MobileNetV2, with fewer learning parameters than the EfficientNet that composed the ESNet.

Cruz-Vega [7] has fewer learning parameters than the proposed DF; however, Cruz-Vega shows a poorer performance. The lower the learning parameters, the faster the computation process. Due to the availability of computer resources, the number of learning parameters is no longer an issue. Despite having fewer learning parameters compared with other CNN models, the proposed DF and ESNet achieve better performances. Here, it

was shown that the accuracy is not determined by the number of learning parameters. The proposed model results in a higher accuracy rate (1.0) compared with that of ResNet (0.9). Similarly, the proposed model achieves a higher accuracy value (1.0) compared with those of the conventional machine learning (0.88). Based on the comparison result, the proposed model achieved the best accuracy result (1.0) with only 3.08 M learning parameters. Only ESNet achieve a similar accuracy of 63%.

**Table 3.** Comparison performance of the proposed DF with other methods. Bold fonts indicate considerable results.

Deep Learning Based					
Methods	Accuracy	Recall	Precision	F-Measure	Learning Parameters
The proposed model	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>3.08 M</b>
ESNet	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	4.88 M
Cruz-Vega [7]	0.820	0.550	<b>1.000</b>	0.860	0.12 M
Khan [48]	0.770	0.920	0.790	0.850	4.90 M
Resnet101	0.90	0.934	0.875	0.904	44.6 M
MS+MP	0.770	0.920	0.790	0.850	3.08 M
Khandakar [14]	0.9581	0.9581	0.9584	0.9582	2.30 M
Conventional Machine Learning Based					
LBP+SVM	0.880	0.917	0.917	0.917	-
LBP+NB	0.710	0.890	0.670	0.760	-
LBP+RF	0.880	0.917	0.917	0.917	-
AdaBoost + Random Forest [14]	0.9671	0.9795	0.9755	0.9775	-

### 5.6. Discussion

The simulations demonstrated that a single CNN model struggles to cope with fewer training images, and this issue can be solved by using the fusion strategy. Fewer training images produce a dataset consisting of low-variance images, which results in highly fluctuating learning curves. The curves also indicate that the trained models might be under-fitting models. Fusing the decisions of two models makes the classification task easier, proven by the maximum accuracy value that it achieves.

The proposed decision fusion model was developed based on the results of the sensitivity and specificity of each model. To recognize the negative patient image, ShuffleNet is activated due to its ability to recognize positive patients well, and vice versa for MobileNetV2. Therefore, the accuracy can reach a score of 1.00. Based on the activation mapping in Figure 19, it was confirmed that fusing MobileNetV2 and ShuffleNet resulted in advantages for DFU classification because both models represent positive and negative images. The proposed fusion strategy resulted in a model with the fewest learning parameters, yet with excellent performances across all evaluation metrics. Combining EfficientNet and MobileNetV2 with ShuffleNet resulted in 4.88 M and 3.08 M learning parameters, respectively, which is less than those of the EfficientNetB0 (ImageNet model) or NasNet-mobile, designed specifically for mobile devices, with 5.3 M learning parameters. Therefore, the proposed fusion rule can be used for CNN model development on mobile devices.

The proposed method has an accuracy of 1.0, which means that all DFU patients were classified correctly. In medical applications, it is very important to classify patients correctly, so that the patient's issue can be detected earlier and handled properly by medical experts. The proposed model, which was a decision-rule-based approach, resulted in a high accuracy rate. These results strengthen the contribution of this study for early DFU detection using thermal imaging. Decision fusion can also contribute to early detection cases in other medical imaging fields.

Classification using a single model has not been able to produce a generalized model that performs well on thermal DFU images. This was caused by several factors, includ-

ing limited and unbalanced datasets. In dealing with those challenges, the study used a data augmentation approach, but data augmentation has great potential to be class-dependent [50]. Another challenge is the dataset availability. There is only one public dataset which was used in this study, whereas training a CNN requires a large and heterogeneous dataset to produce a well-generalized model. To improve the DFU classification performance, it was suggested to increase the number of datasets in the training stage. It would very helpful to increase model performances especially on single CNN models. It also can improve a model ability to classify unseen or blind DFU thermal patient images.

## 6. Conclusions

In this article, we proposed a novel framework to classify the thermal images of diabetic foot ulcers (DFU) based on the fusion schema of two CNN classification results. The CNNs used were MobileNetV2 and ShuffleNet. We applied decision fusion on the CNN classification results and dual-stage validation. Decision-fusion-based classification provides information of the negative and positive labels, which improves the generalization model. ShuffleNet is designed to detect the positive images while MobileNetV2 is used to recognize the negative images. The proposed framework (DF) achieves the maximum values on all performance metrics, namely, accuracy, recall, specificity, precision, and F-Measure. Fusing MobileNetV2 and ShuffleNet also results in a model with fewer learning parameters and a smaller model size. Fusion has shown to be a good solution to achieve a higher performance with a smaller model size. In a future study, enriching the dataset to improve generalization and exploring more fusion strategies to develop smaller CNN models, suitable to be embedded on a mobile device, are potential prospects.

**Author Contributions:** Conceptualization, K.M. (Khairul Munadi), K.S., R.M., M.S., T.F.A. and F.A.; data curation, K.S., M.O. and R.R.; formal analysis, K.M. (Khairul Munadi), K.S., R.M., K.M. (Kahlil Muchtar) and F.A.; funding acquisition, K.M. (Khairul Munadi), R.R. and F.A.; investigation, K.M. (Khairul Munadi), K.S., M.O., R.R., K.M. (Kahlil Muchtar), M.M., R.M., M.S., T.F.A. and F.A.; supervision, K.M. (Khairul Munadi), R.M., M.S. and F.A.; writing—original draft, K.M. (Khairul Munadi), K.S. and F.A.; writing—review and editing, K.M. (Khairul Munadi), K.S., M.O., R.R., K.M. (Kahlil Muchtar), M.M., R.M., M.S., T.F.A. and F.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Education, Culture, Research, and Technology, the Republic of Indonesia, under the 2021 World Class Research Scheme.

**Institutional Review Board Statement:** Not Applicable.

**Informed Consent Statement:** Not Applicable.

**Data Availability Statement:** We did not report any data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* **2009**, *32*, S62–S67.
2. Saminathan, J.; Sasikala, M.; Narayananamurthy, V.; Rajesh, K.; Arvind, R. Computer aided detection of diabetic foot ulcer using asymmetry analysis of texture and temperature features. *Infrared Phys. Technol.* **2020**, *105*, 103219.
3. Usharani, R.; Shanthini, A. Neuropathic complications: Type II diabetes mellitus and other risky parameters using machine learning algorithms. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*. <https://doi.org/10.1007/s12652-021-02972-w>.
4. Ming, A.; Walter, I.; Alhajjar, A.; Leuckert, M.; Mertens, P.R. Study protocol for a randomized controlled trial to test for preventive effects of diabetic foot ulceration by telemedicine that includes sensor-equipped insoles combined with photo documentation. *Trials* **2019**, *20*, 1–12.
5. Wijsinghe, L.; Gamage, C.; Perera, I.; Chitraranjan, C. A smart telemedicine system with deep learning to manage diabetic retinopathy and foot ulcers. In Proceedings of the 2019 Moratuwa Engineering Research Conference (MERCon), Moratuwa, Sri Lanka, 3–5 July 2019; pp. 686–691.
6. Tattersall, G.J. Infrared thermography: A non-invasive window into thermal physiology. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* **2016**, *202*, 78–98.

7. Cruz-Vega, I.; Hernandez-Contreras, D.; Peregrina-Barreto, H.; Rangel-Magdaleno, J.d.J.; Ramirez-Cortes, J.M. Deep Learning Classification for Diabetic Foot Thermograms. *Sensors* **2020**, *20*, 1762.
8. Goyal, M.; Reeves, N.D.; Davison, A.K.; Rajbhandari, S.; Spragg, J.; Yap, M.H. Dfunet: Convolutional neural networks for diabetic foot ulcer classification. *IEEE Trans. Emerg. Top. Comput. Intell.* **2018**, *4*, 728–739.
9. Vardasca, R.; Magalhaes, C.; Seixas, A.; Carvalho, R.; Mendes, J. Diabetic foot monitoring using dynamic thermography and AI classifiers. In Proceedings of the 3rd Quantitative InfraRed Thermography Asia Conference (QIRT Asia 2019), Tokyo, Japan, 1–5 July 2019; pp. 1–5.
10. Alzubaidi, L.; Fadhel, M.A.; Oleiwi, S.R.; Al-Shamma, O.; Zhang, J. DFU\_QUTNet: Diabetic foot ulcer classification using novel deep convolutional neural network. *Multimed. Tools Appl.* **2020**, *79*, 15655–15677.
11. Gamage, C.; Wijesinghe, I.; Perera, I. Automatic Scoring of Diabetic Foot Ulcers through Deep CNN Based Feature Extraction with Low Rank Matrix Factorization. In Proceedings of the 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE), Athens, Greece, 28–30 October 2019; pp. 352–356.
12. Rania, N.; Douzi, H.; Yves, L.; Sylvie, T. Semantic Segmentation of Diabetic Foot Ulcer Images: Dealing with Small Dataset in DL Approaches. In Proceedings of the International Conference on Image and Signal Processing, Marrakesh, Morocco, 4–6 June 2020; Springer Nature Switzerland AG: Berlin/Heidelberg, Germany, 2020; pp. 162–169.
13. Goyal, M.; Yap, M.H.; Reeves, N.D.; Rajbhandari, S.; Spragg, J. Fully convolutional networks for diabetic foot ulcer segmentation. In Proceedings of the 2017 IEEE international conference on systems, man, and cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017; pp. 618–623.
14. Khandakar, A.; Chowdhury, M.E.; Reaz, M.B.I.; Ali, S.H.M.; Hasan, M.A.; Kiranyaz, S.; Rahman, T.; Alfkey, R.; Bakar, A.A.A.; Malik, R.A. A machine learning model for early detection of diabetic foot using thermogram images. *Comput. Biol. Med.* **2021**, *137*, 104838.
15. Mangai, U.G.; Samanta, S.; Das, S.; Chowdhury, P.R. A survey of decision fusion and feature fusion strategies for pattern classification. *IETE Tech. Rev.* **2010**, *27*, 293–307.
16. Faust, O.; Acharya, U.R.; Ng, E.; Hong, T.J.; Yu, W. Application of infrared thermography in computer aided diagnosis. *Infrared Phys. Technol.* **2014**, *66*, 160–175.
17. Suissa, S.; Ernst, P. Optical illusions from visual data analysis: Example of the New Zealand asthma mortality epidemic. *J. Clin. Epidemiol.* **1997**, *50*, 1079–1088.
18. Adam, M.; Ng, E.Y.; Tan, J.H.; Heng, M.L.; Tong, J.W.; Acharya, U.R. Computer aided diagnosis of diabetic foot using infrared thermography: A review. *Comput. Biol. Med.* **2017**, *91*, 326–336.
19. Roslidar, R.; Rahman, A.; Muharar, R.; Syahputra, M.R.; Arnia, F.; Syukri, M.; Pradhan, B.; Munadi, K. A review on recent progress in thermal imaging and deep learning approaches for breast cancer detection. *IEEE Access* **2020**, *8*, 116176–116194.
20. Roslidar, R.; Syaryadhi, M.; Saddami, K.; Pradhan, B.; Arnia, F.; Syukri, M.; Munadi, K. BreaCNet: A high-accuracy breast thermogram classifier based on mobile convolutional neural network. *Math. Biosci. Eng.* **2022**, *19*, 1304–1331.
21. Hernandez-Contreras, D.A.; Peregrina-Barreto, H.; de Jesus Rangel-Magdaleno, J.; Renero-Carrillo, F.J. Plantar thermogram database for the study of diabetic foot complications. *IEEE Access* **2019**, *7*, 161296–161307.
22. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
23. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
24. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
25. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
26. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4510–4520.
27. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6848–6856.
28. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
29. Ioannou, Y.; Robertson, D.; Cipolla, R.; Criminisi, A. Deep roots: Improving cnn efficiency with hierarchical filter groups. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1231–1240.
30. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
31. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.

32. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Baltimore, MD, USA, 17–23 July 2019; pp. 6105–6114.
33. Roggen, D.; Tröster, G.; Bulling, A. Signal processing technologies for activity-aware smart textiles. In *Multidisciplinary Know-How for Smart-Textiles Developers*; Elsevier: Amsterdam, The Netherlands, 2013; pp. 329–365.
34. Oszust, M. Decision fusion for image quality assessment using an optimization approach. *IEEE Signal Process. Lett.* **2015**, *23*, 65–69.
35. Zhang, C.; Sargent, I.; Pan, X.; Gardiner, A.; Hare, J.; Atkinson, P.M. VPRS-based regional decision fusion of CNN and MRF classifications for very fine resolution remotely sensed images. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4507–4521.
36. Kuppusamy, K.; Eswaran, C. Convolutional and Deep Neural Networks based techniques for extracting the age-relevant features of the speaker. *J. Ambient. Intell. Humaniz. Comput.* **2021**, *12*. <https://doi.org/10.1007/s12652-021-03238-1>.
37. James, A.P.; Dasarathy, B.V. Medical image fusion: A survey of the state of the art. *Inf. Fusion* **2014**, *19*, 4–19.
38. Abdi, G.; Samadzadegan, F.; Reinartz, P. Deep learning decision fusion for the classification of urban remote sensing data. *J. Appl. Remote Sens.* **2018**, *12*, 016038.
39. Lai, Z.; Deng, H. Medical image classification based on deep features extracted by deep model and statistic feature fusion with multilayer perceptron. *Comput. Intell. Neurosci.* **2018**, *2018*, 2061516.
40. Rwigema, J.; Mfitumukiza, J.; Tae-Yong, K. A hybrid approach of neural networks for age and gender classification through decision fusion. *Biomed. Signal Process. Control.* **2021**, *66*, 102459.
41. Zhou, T.; Ruan, S.; Canu, S. A review: Deep learning for medical image segmentation using multi-modality fusion. *Array* **2019**, *3*, 100004.
42. Roslidar, R.; Saddam, K.; Arnia, F.; Syukri, M.; Munadi, K. A study of fine-tuning CNN models based on thermal imaging for breast cancer classification. In Proceedings of the 2019 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), Banda Aceh, Indonesia, 22–24 August 2019; pp. 77–81.
43. Janssens, O.; Van de Walle, R.; Locufier, M.; Van Hoecke, S. Deep learning for infrared thermal image based machine health monitoring. *IEEE/ASME Trans. Mechatron.* **2017**, *23*, 151–159.
44. Rizka, R.; Saddam, K.; Roslidar, R.; Munadi, K.; Fitri, A. On Reducing the ShuffleNet Block for Mobile-based Breast Cancer Detection Using Thermogram: Performance Evaluation. *Appl. Comput. Intell. Soft Comput.* **2021**, *under review*.
45. van Netten, J.J.; van Baal, J.G.; Liu, C.; van Der Heijden, F.; Bus, S.A. Infrared thermal imaging for automated detection of diabetic foot complications. *J. Diabetes Sci. Technol.* **2013**, *7*, 1122–1129.
46. Pereira, S.; Pinto, A.; Alves, V.; Silva, C.A. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans. Med. Imaging* **2016**, *35*, 1240–1251.
47. Dai, Y.; Gao, Y.; Liu, F. TransMed: Transformers Advance Multi-Modal Medical Image Classification. *Diagnostics* **2021**, *11*, 1384.
48. Khan, S.; Islam, N.; Jan, Z.; Din, I.U.; Rodrigues, J.J.C. A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. *Pattern Recognit. Lett.* **2019**, *125*, 1–6.
49. Fan, R.E.; Chen, P.H.; Lin, C.J. Working set selection using second order information for training support vector machines. *J. Mach. Learn. Res.* **2005**, *6*, 1889–1918.
50. Balestriero, R.; Bottou, L.; LeCun, Y. The effects of regularization and data augmentation are class dependent. *arXiv* **2022**, arXiv:2204.03632.