WILEY | Hindawi

## Research Article

# Remote Diagnosis and Triaging Model for Skin Cancer Using EfficientNet and Extreme Gradient Boosting

**Irfan Ullah Khan** [ID],[1] **Nida Aslam** [ID],[1] **Talha Anwar** [ID],[2] **Sumayh S. Aljameel** [ID],[1] **Mohib Ullah** [ID],[3] **Rafiullah Khan** [ID],[3] **Abdul Rehman** [ID],[4] and **Nadeem Akhtar** [ID][5]

[1]Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia
[2]National University of Computer and Emerging Sciences, Islamabad, Pakistan
[3]Institute of Computer Science and Information Technology, The University of Agriculture, Peshawar, Pakistan
[4]Department of Computer Science and IT, Virtual University of Pakistan, Lahore, Pakistan
[5]Department of Computer Science and IT, The Islamia University of Bahawalpur, Bahawalpur, Pakistan

Correspondence should be addressed to Nida Aslam; naslam@iau.edu.sa

Due to the successful application of machine learning techniques in several fields, automated diagnosis system in healthcare has been increasing at a high rate. The aim of the study is to propose an automated skin cancer diagnosis and triaging model and to explore the impact of integrating the clinical features in the diagnosis and enhance the outcomes achieved by the literature study. We used an ensemble-learning framework, consisting of the EfficientNetB3 deep learning model for skin lesion analysis and Extreme Gradient Boosting (XGB) for clinical data. The study used PAD-UFES-20 data set consisting of six unbalanced categories of skin cancer. To overcome the data imbalance, we used data augmentation. Experiments were conducted using skin lesion merely and the combination of skin lesion and clinical data. We found that integration of clinical data with skin lesions enhances automated diagnosis accuracy. Moreover, the proposed model outperformed the results achieved by the previous study for the PAD-UFES-20 data set with an accuracy of 0.78, precision of 0.89, recall of 0.86, and F1 of 0.88. In conclusion, the study provides an improved automated diagnosis system to aid the healthcare professional and patients for skin cancer diagnosis and remote triaging.

## 1. Introduction

Skin cancer is one of the commonly occurring and deadly types of cancer. The expected estimated number of newly diagnosed skin cancer patients during 2020 in the USA will be more than 1.8 million [1]. Skin cells are usually damaged due to excessive exposure to ultraviolet (UV) radiation. Skin cancer is a type of cancer caused by damaged skin cells or abnormal growth of skin cells. It can be mainly categorized as basal cell carcinoma (BCC), melanoma (MEL), non-melanoma skin cancer, and squamous cell carcinoma (SCC). However, some types of skin cancer are very rare such as Kaposi sarcoma (KS) and actinic keratosis (AK), also known as solar keratosis, lymphoma, and keratoacanthoma. The nature of some skin cancer types is lethal and metastasize in nature. Early screening and prognosis of the skin cancer will increase the chance of recovery and survival; otherwise, it will lead to grim conditions.

The widespread and deadly nature of the disease demands the effective noninvasive diagnostic mechanism with increased accuracy. Skin cancer is mainly diagnosed via visual examination along with some clinical and histological investigations. The clinical information includes some demographic information, location, and nature of skin lesion [2]. Visual examination with the naked eye usually could not recognize and disclose the details. To overcome this

drawback, dermatoscope, a medical equipment for the skin lesion investigation, was introduced. The device greatly enhanced the accuracy of early diagnosis capability [3]. A dermoscopic image is a magnified high-resolution enlarged image of the skin lesion.

Despite the invention of dermoscopic images, which greatly enhanced the accuracy, it highly depends upon the dermatologist's experience and subjective judgment [3]. High similarity among the visual feature of different types of skin cancer sometimes leads to the wrong diagnosis. The diagnosis can be further enhanced using seven checklist points [4] and ABCD and ABCDE rules [5]. Seven checklist points use seven dermoscopic features for malignant melanoma diagnosis. The sensitivity of the diagnosis was further improved with the integration of dermoscopic images and seven checklist point [6]. While, in the ABCDE rule, A stands for asymmetry; B stands for border; C stands for colour; D stands for differential structure; and E stands for evolution [7]. The ABCDE rule increases the accuracy of the diagnosis but also require proper training to use the criteria. Previously mentioned methods improved the diagnosis process but were limited to melanoma diagnosis only. Therefore, demand for an automated computationally intelligent method that can further enhance the visual features and can aid the dermatologist in the diagnosis.

The rest of the paper is organized as follows. Section 2 reviews previous studies for skin cancer diagnosis. Section 3 presents the material and methods used in the study. Section 4 provides the experimental setup and results. Section 5 contains the conclusion.

## 2. Related Studies

Several studies have been made to develop a computer-aided diagnosis (CAD) system for skin cancer [8, 9]. Initially, the studies related to skin cancer were mainly focused on using image processing techniques [10], following using machine learning techniques (supervised and unsupervised) [11] and recently convolution neural network (CNN) and deep learning models [12]. Deep learning models have produced significant development and advancement in medical image analysis and particularly for skin cancer [13, 14]. Some of the recent studies using deep learning are discussed below. The studies are organized chronologically in the literature review.

Esteva et al. used the deep CNN model for the diagnosis of two types of skin cancer such as keratinocyte carcinomas and seborrheic keratoses. The diagnosis performance was compared against the decision made by 21 highly qualified and experienced dermatologists [15]. The study proved the significance of AI and particularly deep learning in skin cancer diagnosis. The study was performed using ISIC and Dermofit skin lesion data set. Another investigation was made by Haenssle et al. [16] who compared the performance of the Google Inception V4 deep learning model with the top five algorithms in the ISIC 2016 challenge and the diagnosis decision made by 58 dermatologists. The data set contains both the images (dermoscopic and digitalized images) and clinical information of 100 patients. Furthermore, a study made by Brinker et al. [17] developed an enhanced deep learning model

for the diagnosis of melanoma. They compared the performance of the model with the diagnosis decision made by 145 dermatologists from 12 hospitals in Germany.

Additionally, Pacheco et al. [18] developed a smartphone application using skin lesion images and clinical information for automated diagnosis. The study covers six categories of skin cancer with a total of 1,641 skin lesions. The study compared various deep learning models such as GoogleNet, VGGNet-13/19-bn, ResNet (50, 101), MobileNet, and three-layer convolutional neural network. The models were first trained using skin lesion images taken from smartphone cameras and then using both (skin lesion images and clinical features). The first model achieved an accuracy of 0.69 and was enhanced with the integration of clinical data and achieved an average accuracy of 0.764. The proposed study attempts to enhance the outcome achieved by Pacheco's study.

Consequently, Kadampur and Riyaee's study developed a model-driven architecture for the diagnosis of skin cancer using dermal cell images. Several deep learning models were trained using the HAM10000 data set and achieved an AUC (area under the curve) of 0.99 [19]. Likewise, two CNN models such as region-based convolutional neural network (RCNN) and Faster RCNN was used for skin lesion classification for benign and malignant tumor images [20]. The outcome of the model was compared with the diagnosis made by 10 certified dermatologists and 10 trainee dermatologists and conclusively achieved better classification accuracy than the dermatologists.

Wei et al. [21] used pretrained deep learning models such as MobileNet and DenseNet. ImageNet weights were used for extracting the features using ISIC 2016 data set and achieved an accuracy of 0.962. Another study performed by Pham et al. [22] developed a CNN model for melanoma classification using ISIC 2019 and MClass-D set dermoscopic skin lesion and achieved an AUC of 0.944. The diagnosis of the proposed framework was further verified with the 157 certified dermatologists in German hospitals.

Importantly, the integration of skin cancer clinical images and intelligent computation techniques produced effective outcomes and motivated the exploration and implication of remote triaging for the skin cancer diagnosis. Recently, Udrea et al. [23] proposed smartphone application for identifying the patient at risk using a skin lesion. The model was trained using skin lesion images from multiple data sets. Initially, the lesion segmentation was applied on the image, after the segmentation noise such as hairs and freckles were removed; then features were extracted such as colour, shape, and texture; and finally all the extracted features were input to support vector machine (SVM) classifier. The application produced good outcomes in terms of sensitivity (0.951).

The transfer learning concept was widely used for skin cancer detection. One of the studies performed by Kassem et al. [24] on the ISIC 2019 challenge data set using the GoogleNet pretrained model for eight categories of skin cancer lesion and achieved an accuracy of 0.949. Another extensive study was made to evaluate the performance of the proposed YOLOv2-SquuezeNet for segmentation and

several classifiers for classification using four-year ISIC data sets challenges (2017, 2018, 2019, and 2020) [25]. The study achieved mean average precision of 985 using optimized SVM. Moreover, Gessert et al. [26] proposed an ensemble method for integrating gender, anatomy information, and skin lesion for diagnosing skin cancer using multiple data sets. Several image processing techniques were applied for preprocessing. However, the model was trained using EfficientNet and achieved an accuracy of 0.63 using the ISCIS 2019 data set.

Recently a study made by Goceri [27] developed a multilayered deep learning model using facial skin lesions. Initially, the images were segmented to identify the facial disorder skin lesion. Subsequently, these skin lesions were used by pretrained DenseNet201 for classifying the skin lesions. The study achieved an accuracy of 0.95. Furthermore, another study was performed for melanoma diagnosis using dermoscopic images [28]. ISIC 2020 data set was used for training EfficientNet models (B5 and B6). The study extensively applied several data augmentation techniques to increase the number of images and better training the deep learning models. They achieved an accuracy of 0.9411.

Despite extensive research made in skin cancer diagnosis, mostly, studies are using skin lesion images, and very few studies used clinical data. The importance of clinical data in the diagnosis cannot be denied [15, 16, 22]. One of the recent studies made by Pacheco and Krohling used the data set consist of digitalized images taken by smartphone cameras along with the clinical data of the patients [18]. The data set covers multiple types of skin cancer. Regardless of significant results made by the researcher for the specified data set, the results can be further enhanced, and several techniques can be used and integrate to better train the model.

## 3. Materials and Methods

This section contains the description of the data set (PAD-UFES-20) used in the studies, data preprocessing, and classification model used in the study.

*3.1. Data Set Description.* PAD-UFES-20 [29] data set was collected under the Dermatological and Surgical Assistance Program (PAD) at Federal University of Espírito Santo. The PAD-UFES-20 data set consists of skin lesions and clinical data with an average patient's age of 60 years. The data set contains the data of 1,373 patients, 1,641 skin lesions, 2,298 images, and metadata containing 26 attributes. Some of the images were removed due to the low-quality phone camera used to capture the image. Some of the patients have more than one type of skin cancer lesion. The number of images per category in the data set is shown in Figure 1. PAD-UFES-20 suffers from class imbalance; the number of images for ACK and BCC is high when compared with other categories. The data set covers six types of skin cancer such as actinic keratosis (ACK), basal cell carcinoma (BCC), melanoma (MEL), nevus (NEV), squamous cell carcinoma (SCC), and seborrheic keratosis (SEK).

Moreover, the data set contains metadata, that is, clinical features (26 attributes) in addition to the skin lesions. Some of the attributes are the identifiers and were removed; 21 features are clinical data and the class label. Some of the features are demographic information such as age, smoking and drinking habits, and father and mother background. Some of the attributes are related to the lesions such as itches, bleed, hurt, and so on. Clinical features are established on the questions commonly asked by dermatologists. The description of the vital signs attributes in data set is shown in Table 1.

Some of the skin cancer have common regions in the human body. For example, SEK skin cancer type lesion is more common in the face region; however, ACK is common in the forearm, and NEV is more common in the back region. The occurrence of skin cancer lesions in different body regions is shown in Figure 2. Similarly, ACK, BCC, and MEL skin cancer types do not grow. However, SEK category has equal distribution of lesion that sometimes grow and sometimes does not grow. Figure 3 shows the distribution of skin lesions per category based on the attribute grew. ACK, BCC, and SCC types of skin moles are itchy in nature. Figure 4 shows the itchiness' nature of the different types of skin cancer. However, only ACK type of skin cancer hurts and bleeds when compared with the other five types. Some of the sample images from PAD-UFES-20 for each category of skin cancer are presented in Figure 5.

Most of the features in the data set are categorical except age, Fitzpatrick, diameter_1, and diameter_2. The statistical description of the numerical features is presented in Table 2. The prevalence of some types of skin cancers such as ACK, BCC, MEL, SCC, and SEK is in the age range of 59.9 to 68.86 years. However, the mean ($\mu$) age for the NEV category is 35.64. The minimum age of the patients in the data set is 6 years, and the maximum age is 94 years. Similarly, the mean of diameter_1 for BCC, NEV, and SCC is similar. However, the mean of the diameter_2 is similar for ACK, BCC, and SCC categories.

*3.2. Data Preprocessing and Augmentation.* For better generalization of the deep learning model and to alleviate the data imbalance problem, data augmentation technique was applied. The data imbalance usually led to model overfitting for the majority class. Augmentation was applied via resizing, flipping, shifting, and rotating. For resizing the images, the zoom range of 0.1 and rescale of 1.0/255 were set, while a dimension of $300 \times 300 \times 3$ was used, which is a recommended input size for EfficientNetB3. Moreover, random horizontal and vertical flipping along with width and height shifting with a range of 0.1 were used to increase the generalization of the model for all possible locations of the skin cancer in images. For some images, 360° rotation was performed. The data augmentations were only applied to the training data set.

*3.3. Classification Model.* After the data augmentation classification model was developed. The proposed model consists of two classification models such as EfficientNet for
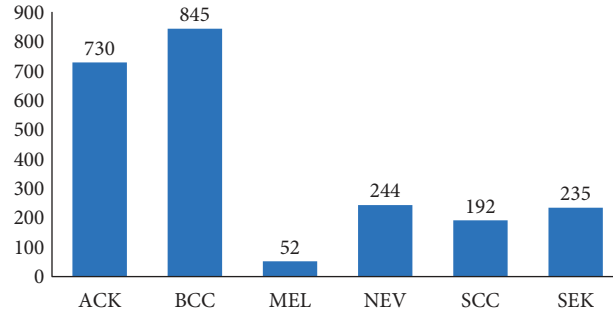
Figure 1: Number of skin cancer lesions per category in the PAD-UFES-20 data set.

Table 1: Description of the PAD-UFES-20 clinical data.

| Features | Data type | Data set description |
|---|---|---|
| Fitzpatrick | Categorical (1–5) | It is a phototype used to define a person's skin type in terms of response to (UVR) exposure |
| Region | Categorical (15 categories) | The location of the skin lesion in the patient's body |
| Diameter_1 | Numerical | Skin lesions' horizontal diameters |
| Diameter_2 | Numerical | Skin lesions' vertical diameters |
| Diagnostic | Categorical (6 categories) | Skin cancer types, that is, ACK, MEL, NEV, and so on |
| Itch | Boolean | Skin lesion itching |
| Grew | Boolean | Skin lesion grew |
| Hurt | Boolean | Skin lesion hurts |
| Changed | Boolean | Skin lesion has recently changed |
| Bleed | Boolean | Skin lesion is bleeding |
| Elevation | Boolean | Skin lesion has an elevation |
| Biopsed | Boolean | Diagnosis from clinical data or lab test biopsy |
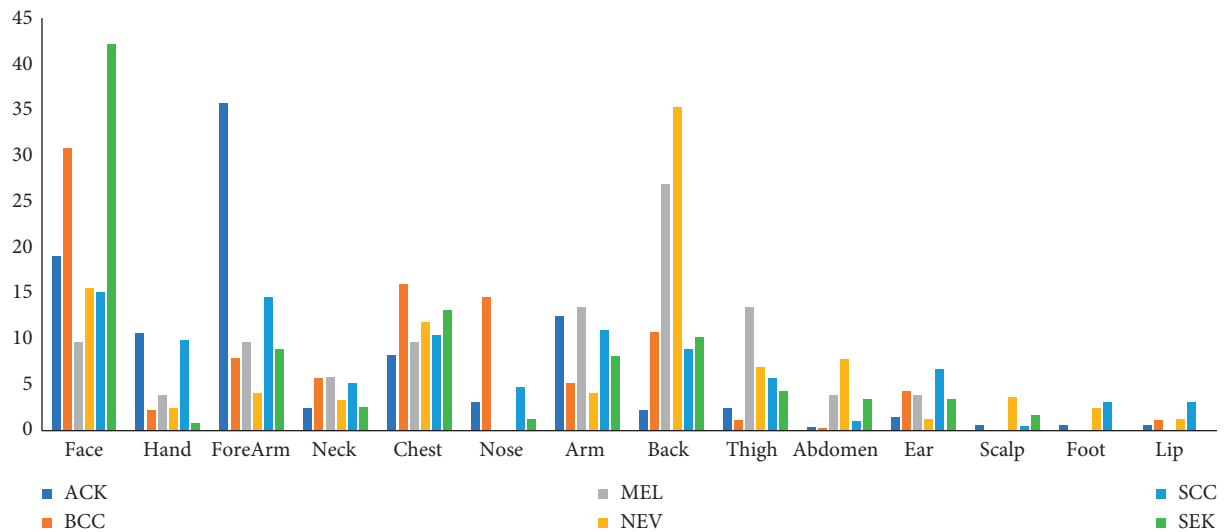


Figure 2: Presence of skin cancer lesion per category in different body regions.

skin lesions and Extreme Gradient Boosting (XGB) for the clinical data. The description of both models is discussed below.

### 3.4. EfficientNet Deep Learning Model.
Deep learning (DL) is a kind of convolutional neural network (CNN) and is widely used for images [30]. Recently deep learning has been widely used for the diagnosis of various medical diseases. Similarly, some studies have been made on diagnosis of skin diseases using deep learning [12]. DL consists of multiple connected layers using various weights and activation functions. The basic deep learning model contains a convolutional layer, pooling, and connected layers. Several activation functions are used to adjust the weights. The activation functions create a feature map that is input into the subsequent layer.
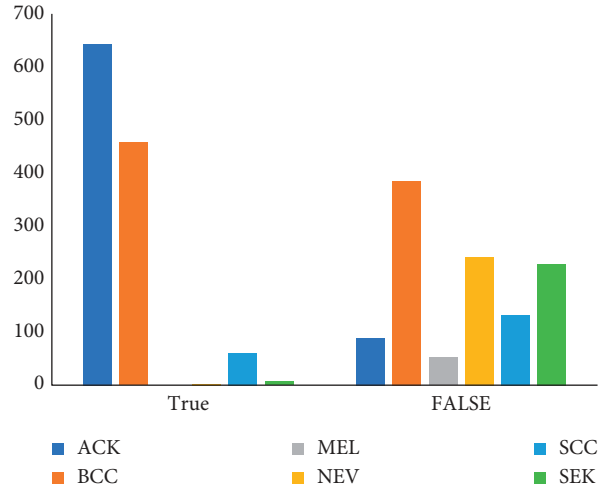
FIGURE 3: Percentage of skin lesion per skin cancer category for grew attribute.
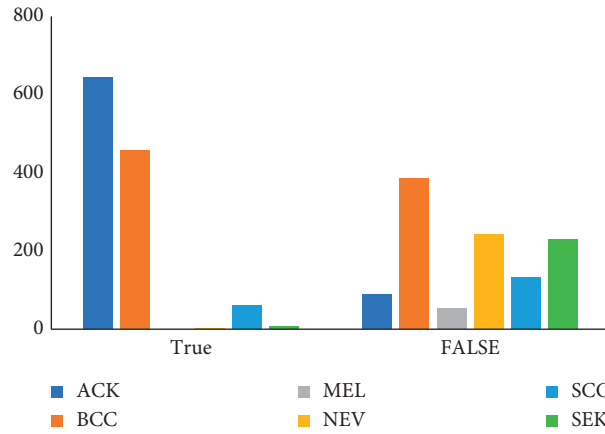


FIGURE 4: Percentage of skin lesion per skin cancer category for itch attribute.

Pooling and convolutional layers are used for extracting the features. These layers are used for extracting the visual features and understand the complex nature of the images. Nevertheless, the nature of the skin cancer lesion is very complex, and developing an automated diagnosis system using deep learning is challenging. To alleviate this problem, transfer learning is used.

In our study, EfficientNetB3 is used for skin cancer detection. EfficientNetB3 is an up-to-date, cost-efficient, and robust model developed by scaling three parameters such as depth, width, and resolution [31]. The EfficientNetB3 model with noisy-student weights is used in scenarios I and III for the transfer learning process, while "isicall_eff3_weights" weights are used as pretrain for scenarios II and IV. The GlobalAveragePooling2D layer is added to each scenario to generalize the model better. The number of parameters were reduced. Furthermore, the RELU activation function is used with three dense and two dropout layers. The output layer contains multiple output units for multiclass classification using the softmax activation function. Table 3 enlists layers, parameters, weights, and so on used in the proposed study.

### 3.5. Extreme Gradient Boosting (XGB).

Extreme Gradient Boosting (XGB) is an ensemble-based classification algorithm and was proposed by Chen in 2015 [32]. XGB uses boosted tree and is used for classification and regression. XGB has been widely used for various prediction task and produces significant outcome due to efficient learning capability and speed. XGB is an enhanced version of the gradient boosting tree. The main aim of the algorithm is the optimization of the objective function by reducing the loss, model complexity, and computational resource utilization. The complexity is reduced using regularization. Moreover, the technique normalization is used to alleviate the model overfitting. The aim of using XGB for clinical data is due to its innate capability to handle the data imbalance.

The algorithm works by adding the trees iteratively by splitting the features. In every next iteration, new rules are added, and the loss decreases. The iteration continues until the model achieved the required optimal performance. XGB model uses the second-order derivative to the loss function. Assume $D$ is the data set consists of $n$ number of attributes as follows:

Figure 5: Sample images for each category of skin cancer lesion from the PAD-UFES-20 data set.

TABLE 2: Statistical description of clinical features in the PAD-UFES-20 data set.

| Features | ACK $\mu \pm \sigma$ | BCC $\mu \pm \sigma$ | MEL $\mu \pm \sigma$ | NEV $\mu \pm \sigma$ | SCC $\mu \pm \sigma$ | SEK $\mu \pm \sigma$ |
|---|---|---|---|---|---|---|
| Age | $61.7 \pm 12.7$ | $63.15 \pm 13.46$ | $59.9 \pm 13.9$ | $35.64 \pm 15.59$ | $68.86 \pm 12.21$ | $65.89 \pm 12.2$ |
| Fitzpatrick | $2.33 \pm 0.73$ | $2.2 \pm 0.69$ | $2.29 \pm 0.69$ | $2.7 \pm 0.87$ | $2.21 \pm 0.66$ | $2.7 \pm 1.13$ |
| Diameter_1 | $10.98 \pm 7.1$ | $12.05 \pm 8.12$ | $15.2 \pm 7.60$ | $12.13 \pm 17.6$ | $12.22 \pm 8.41$ | $9.37 \pm 5.1$ |
| Diameter_2 | $8.19 \pm 5.22$ | $9.02 \pm 6.02$ | $11.17 \pm 6.49$ | $7.35 \pm 6.64$ | $9.51 \pm 5.12$ | $6.91 \pm 3.78$ |

$$D = \{x_1, x_2, \ldots, x_n\}. \tag{1}$$

$Y$ represents the class attribute; $Y_i$ represents the actual value, while $Y_t'$ represents the predicted value.

$$\text{Tree\_Ens} = \sum_{k=1}^{j} \text{Loss}\left(y_k, \sum_{n=1}^{N} f_n(x_k)\right) + \sum_{n=1}^{N} \Omega(f_n), \quad f_n \in F, \tag{2}$$

where Tree_Ens represents a tree ensemble model. The loss represents loss function, which is the difference between the predicted and the actual. $N$ represents the number of trees. $F$ represents the set of the trees used in the model training. $\Omega$ represents the regularization term.

## 4. Experimental Setup and Results

The models were implemented using python 3.8.4, and experiments were carried out on Google Colab notebook using GPU run-time type. Experiments were conducted on original and augmented data sets. Different experimental scenarios are discussed in the section below:

*Scenario I.* EfficientNet noisy-student weights (PAD_UFES-20 data set): In this scenario, EfficientNet noisy student's weight was used for training the model. Initially, the weights were computed using ISIC 2019 data set. Later, the model was further trained and tested using PAD-UFES-20 images. The noisy student is a semisupervised technique that enhances the training and purification of the model [33]. It enhanced the performance of the ImageNet. The main idea behind the noisy student is that the number of students is either equal to or greater than the number of teachers, with the aim that the larger the number better will be the training. Secondly, the noise will be added so that noisy students are pushed to learn better and harder from the data set.

*Scenario II.* ISIC 2019 weights (PAD_UFES-20 data set): ISIC 2019 weights were used to train the model [34]. However, PAD_UFES-20 skin lesions were used for training and testing.

*Scenario III.* EfficientNet noisy-student weights (PAD_UFES-20 data set): In this scenario, EfficientNet noisy student's weight was used for training the model. In this scenario, skin lesions and clinical data were also used.

*Scenario IV.* ISIC 2019 weights (PAD_UFES-20 data set): ISIC 2019 weights were used to train the model.

However, PAD_UFES-20 skin lesions and clinical data were used for training and testing.

During the experiments, the stratified fivefold cross-validation method was used. Moreover, 30 epochs with 76 steps per epoch and a batch size of 24 were used. The learning rate was set to 0.0001, and the ReduceLROnPlateau method was used to investigate the validation loss. The study defined factor = 0.5, patience = 5, and the min_lr = 0.000001 in the proposed method. The ADAM with 0.001 optimization method was used as a solver. The model has been validated in two different phases.

The performance of the proposed model is evaluated in terms of several standard evaluation measures such as accuracy, precision, recall, F1 measure, and AUC (area under the curve). The precision of the proposed model is the ratio of skin lesions that are correctly predicted as skin cancer types. However, the recall is the ratio of correctly predicted skin lesions. Similarly, accuracy is correctly predicted skin lesions as skin cancer types. While the F1 measure is the harmonic mean of recall and precision. The values have been computed for each fold and finally, the average values for each fold have been calculated. The below figures show the loss and accuracy for training and validation for all the scenarios. Figures 6(a)–6(d) represents the curve for scenario I; Figures 7(a)–7(d) represents the curve for scenario II; Figures 8(a)–8(d) represents the curve for scenario III; and Figures 9(a)–9(d) represents the curve for scenario IV.

The results of the proposed model are represented in Table 4 for all four scenarios. The average highest accuracy reported by the experiments was same for scenarios III and IV. ISIC 2019 weights have outperformed the noisy student's weights with combined data, that is, skin lesion and clinical data. However, with only skin lesions, noisy students produced a better outcome. The reported average accuracy of 0.76, recall of 0.82, same precision, and F1 and AUC of 0.81. The study confirmed the finding made by Pacheco's study [18] that the integration of clinical data enhances the diagnosis and triaging performance.

The proposed study outcome was compared with the literature study. Indeed, it is important to mention that only one study was used for comparison because so far only one study has used the PAD-UFES-20 data set. The results achieved in the proposed study were not compared with the studies in the literature using ISIC 2019 data set because in the current study, the results were achieved using the PAD-UFES-20 data set.

PAD-UFES-20 data set was proposed and used by Pacheco. The results presented in Table 5 confirmed that the proposed study outperformed [24] in terms of all specified

TABLE 3: Summary of the proposed model.

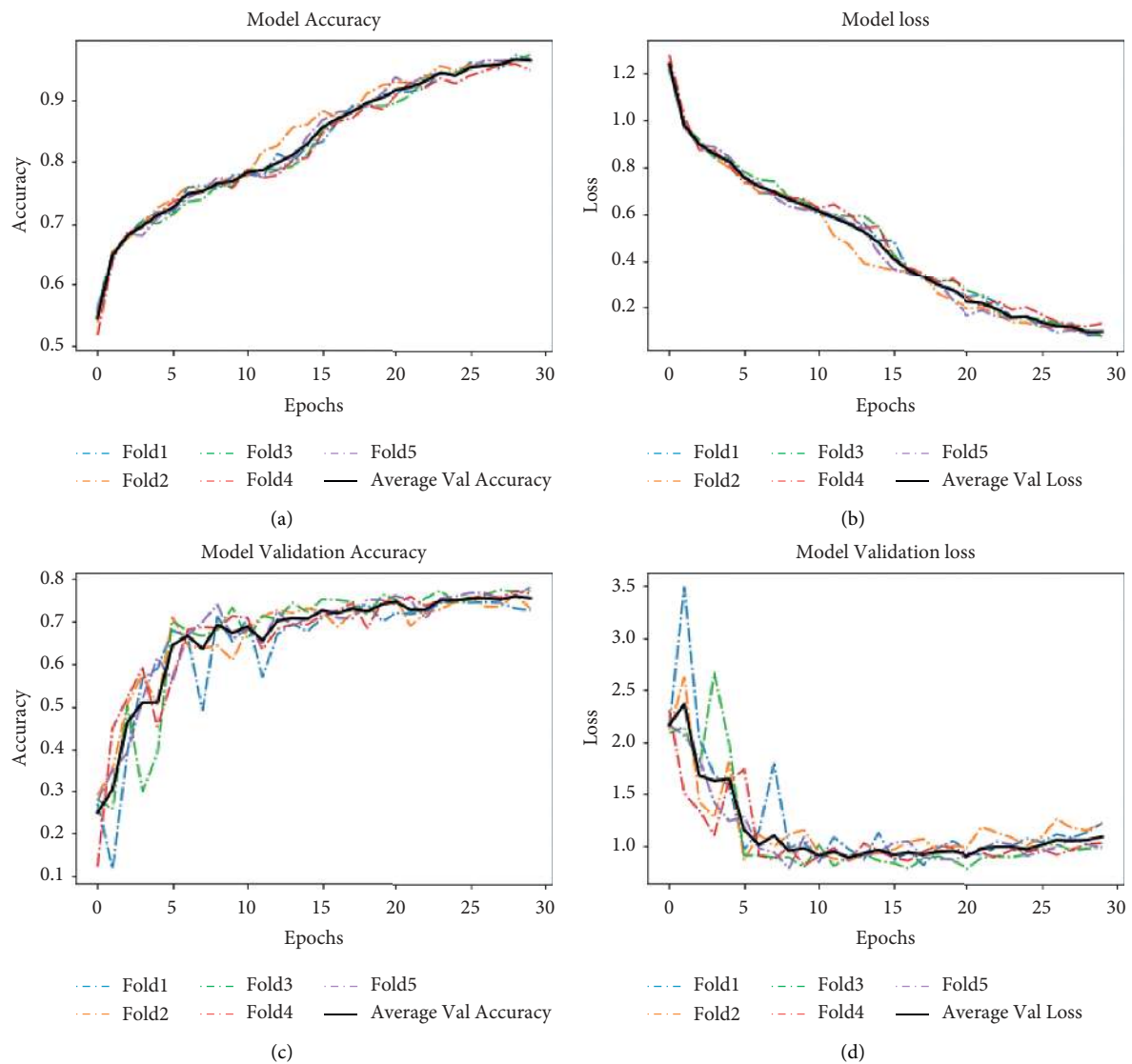| Layer (type) | Output shape | Parameters |
| --- | --- | --- |
| Input layer | 300, 300, 3 | 0 |
| EfficientNetB3 (model) | 10, 10, 1,536 | 10,783,528 |
| Global_average_pooling2D | 1,536 | 0 |
| Dropout_2 | 1,536 | 0 |
| Dense_2 | 128 | 196,736 |
| Dropout_3 | 128 | 0 |
| Dense_3 | 64 | 8,256 |
| Dense_4 | 6 | 390 |
| Total parameters | — | 10,988,910 |
| Trainable parameters | — | 10,901,614 |
| Nontrainable parameters | — | 87, 296 |



FIGURE 6: Scenario I model: (a) training accuracy with number of epochs, (b) training loss with number of epochs, (c) validation accuracy with number of epochs, and (d) validation loss with number of epochs.

measures except AUC. However, similar results have been achieved for both scenarios based on all the selected evaluation measures except AUC for both cases using skin lesion and combined data set (skin lesion and clinical features). Moreover, the AUC achieved by the study in both scenarios is similar.
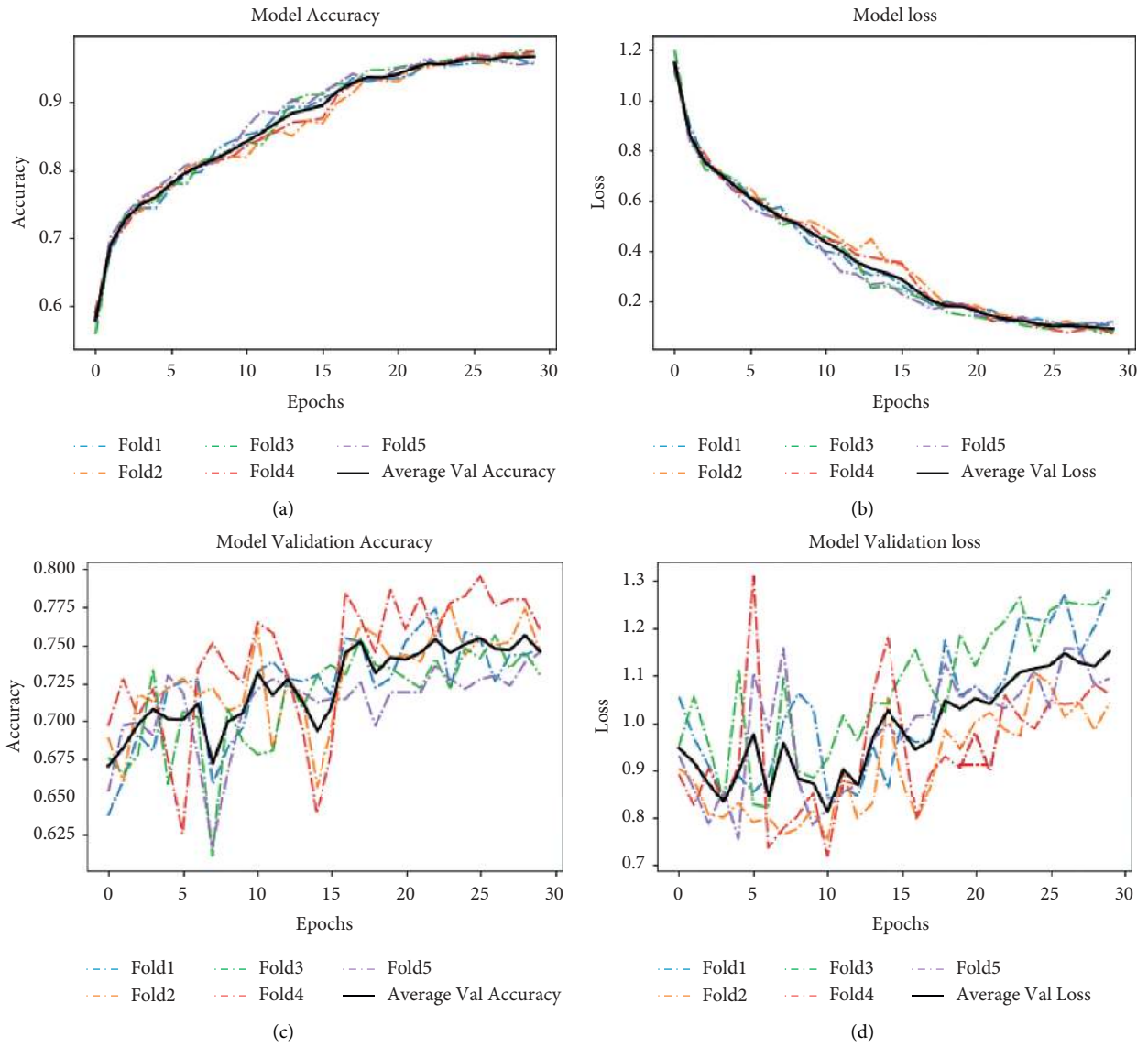
FIGURE 7: Scenario II model: (a) training accuracy with number of epochs, (b) training loss with number of epochs, (c) validation accuracy with number of epochs, and (d) Validation loss with number of epochs.
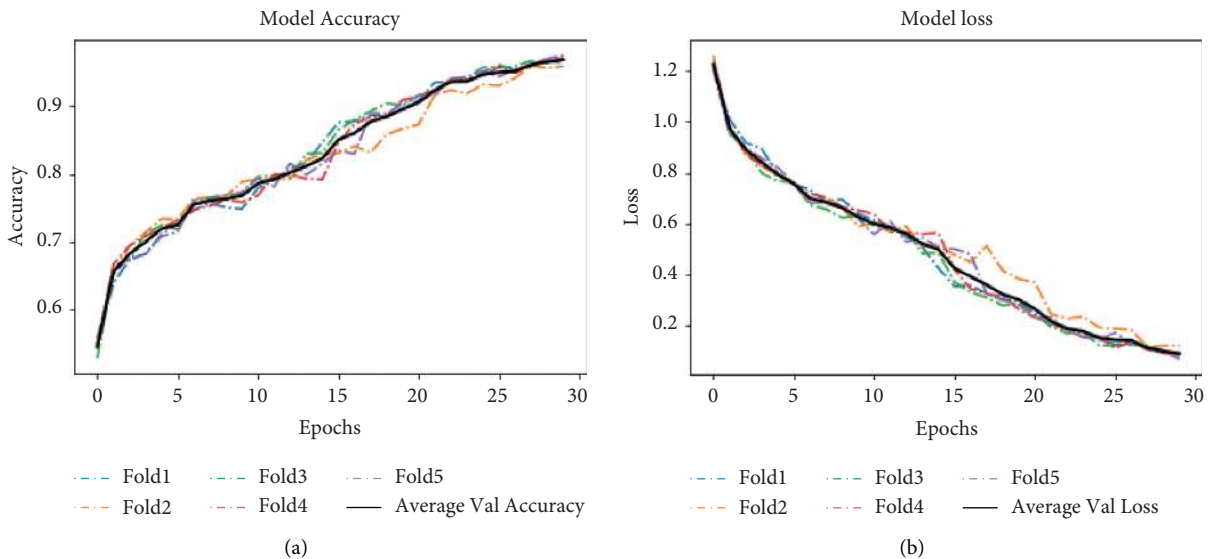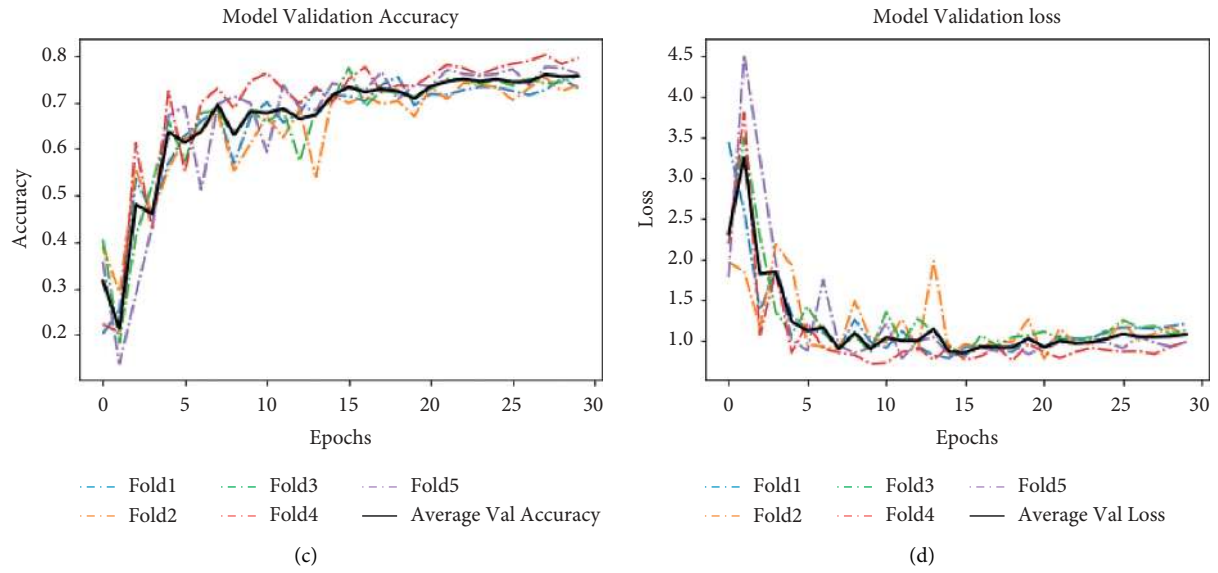


FIGURE 8: Continued.

(c)

(d)

FIGURE 8: Scenario III model: (a) training accuracy with number of epochs, (b) training loss with number of epochs, (c) validation accuracy with number of epochs, and (d) validation loss with number of epochs.
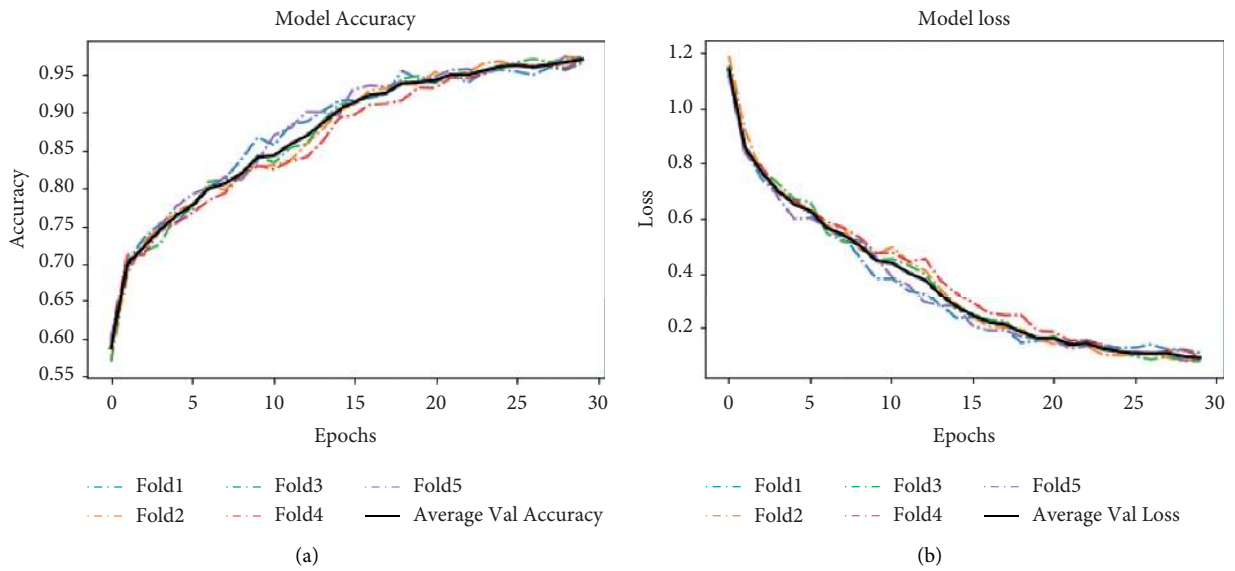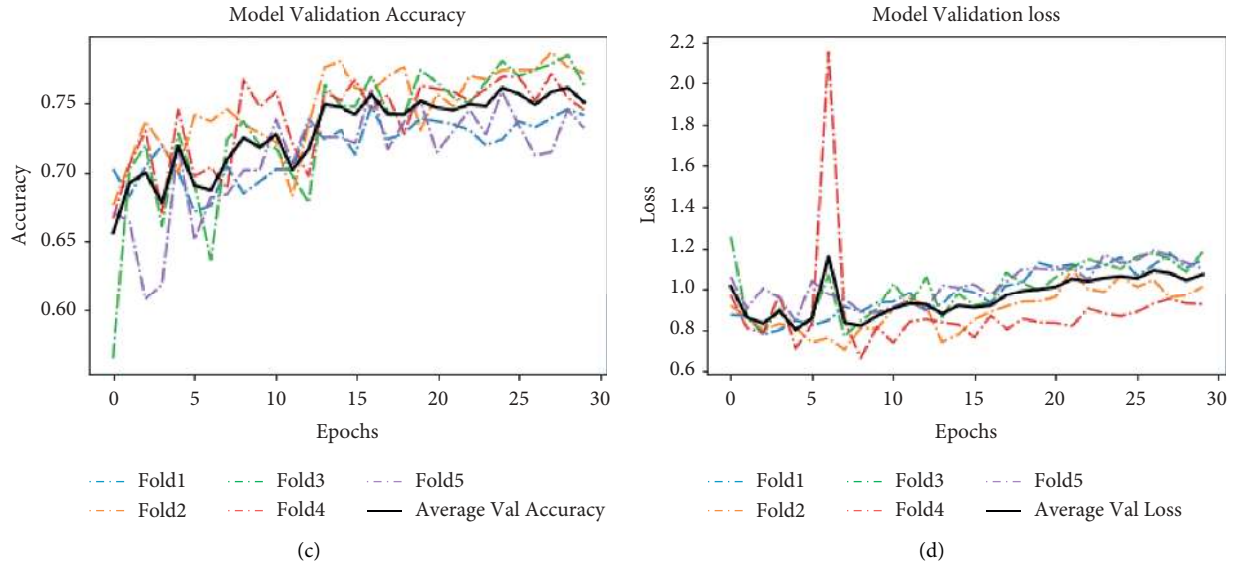


(a)

(b)

FIGURE 9: Continued.

FIGURE 9: Scenario IV model: (a) training accuracy with number of epochs, (b) training loss with number of epochs, (c) validation accuracy with number of epochs, and (d) validation loss with number of epochs.

TABLE 4: Experimental results obtained from the test set using fivefold stratified cross-validation.

| Experimental scenarios | Data type | Accuracy | Precision | Recall | F1 | AUC |
| --- | --- | --- | --- | --- | --- | --- |
| Scenario I | Skin lesion | 0.76 | 0.81 | 0.82 | 0.81 | 0.81 |
| Scenario II | Skin lesion | 0.74 | 0.78 | 0.8 | 0.79 | 0.80 |
| Scenario III | Skin lesion + clinical data | 0.78 | 0.88 | 0.85 | 0.87 | 0.83 |
| Scenario IV | Skin lesion + clinical data | 0.78 | 0.89 | 0.86 | 0.88 | 0.82 |

TABLE 5: Results of the proposed study and comparison with results from Pacheco's study [24].

| Data | Reference | Accuracy | Precision | Recall | F1 | AUC |
| --- | --- | --- | --- | --- | --- | --- |
| Skin lesions | Pacheco study [18] | 0.690 | 0.72 | 0.690 | 0.695 | 0.92 |
| | Proposed study | 0.76 | 0.81 | 0.82 | 0.81 | 0.81 |
| Skin lesions + clinical features | Pacheco study [18] | 0.76 | 0.77 | 0.76 | 0.76 | 0.94 |
| | Proposed study | 0.78 | 0.89 | 0.86 | 0.88 | 0.82 |

Despite the significant results achieved by the proposed study, there is still room for improvement. The study used the data augmentation technique for alleviating data imbalance and better model generalization. Thus, it is recommended to collect more skin lesions for the categories that have very a smaller number of samples as compared to the other categories. Moreover, some of the clinical features were missing. Similarly, some of the images in the data set do not have diagnosis using biopsy; the diagnosis decision was made using the dermatologist diagnosis. Therefore, there is a need for the data set where all the diagnosis for the skin lesions was made using biopsy. However, the proposed study overall produced high results when compared with the original study that proposed the data set.

Conclusively, the main contributions are:

(1) The study explores the impact of clinical data on diagnosis of skin cancer using skin lesions and attempts to propose an automated tool for early diagnosis

(2) For better generalization of the proposed model, data augmentation techniques were applied

(3) In general, the proposed study model has outperformed the benchmark study and can be served as an effective tool for the diagnosis and triaging of skin cancer

## 5. Conclusions

This research presents an automated diagnosis and triaging system for skin cancer. The study used the EfficientNetB3 model for analysing the images taken via smartphone cameras; however, for clinical data, Extreme Gradient Boosting (XGB) ensemble classifier model is used. The main reason for using the XGB classifier is due to its better performance in the imbalanced data sets. The proposed study confirms the findings made by Pacheco's study, that is, the integration of clinical data enhances the diagnosis and triaging performance. The average accuracy reported in the study was 0.76 using skin cancer images and 0.78 using

images and the clinical data. The proposed study outperformed the benchmark study. Despite the data imbalance limitation, data augmentation techniques were applied to reduce the risk of model overfitting. Nevertheless, the outcome was significant, but there is still further need for improvement. The model can be further enhanced by implementing and comparing other deep learning models. Furthermore, the proposed model needs to be tested on multiple data sets. However, there is no other open-source data set available for skin cancer diagnosis that contains the skin cancer lesions and the clinical data.

## Data Availability

The study was performed using PAD-UFES-20 and can be accessed from the web link, https://data.mendeley.com/datasets/zr7vgbcyr2/1.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## References

[1] NIH National Cancer Institute, "Cancer facts and figures 2020," *CA: A Cancer Journal for Clinicians*, vol. 70, pp. 1–76, 2020, https://www.cancer.org/cancer/bladder-cancer/detection-diagnosis-staging/survival-rates.html.

[2] K. Wolff, R. A. Johnson, A. P. Saavedra, and E. K. Roh, *Color Atlas and Synopsis of Clinical Dermatology*, McGraw-Hill Education., New York, NY, USA, 8th edition, 2017.

[3] C. Sinz, P. Tschandl, C. Rosendahl et al., "Accuracy of dermatoscopy for the diagnosis of nonpigmented cancers of the skin," *Journal of the American Academy of Dermatology*, vol. 77, no. 6, pp. 1100–1109, 2017.

[4] M. F. Healsmith, J. F. Bourke, J. E. Osborne, and R. A. Graham-Brown, "An evaluation of the revised seven-point checklist for the early diagnosis of cutaneous malignant melanoma," *British Journal of Dermatology*, vol. 130, no. 1, pp. 48–50, 1994.

[5] F. Nachbar, W. Stolz, T. Merkle et al., "The ABCD rule of dermatoscopy: high prospective value in the diagnosis of doubtful melanocytic skin lesions," *Journal of the American Academy of Dermatology*, vol. 30, no. 4, pp. 551–559, 1994.

[6] V. Dal Pozzo, C. Benelli, and E. Roscetti, "The seven features for melanoma: a new dermoscopic algorithm for the diagnosis of malignant melanoma," *European Journal of Dermatology*, vol. 9, no. 4, pp. 303–308, 1999.

[7] J. K. Robinson and R. Turrisi, "Skills training to learn discrimination of ABCDE Criteria by those at risk of developing melanoma," *Archives of Dermatology*, vol. 142, 2006.

[8] L. Smith and S. Macneil, "State of the art in non-invasive imaging of cutaneous melanoma," *Skin Research and Technology*, vol. 17, no. 3, pp. 257–269, 2011.

[9] A. Masood and A. Ali Al-Jumaily, "Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms," *International Journal of Biomedical Imaging*, vol. 2013, pp. 1–22, 2013.

[10] R. H. Moss, W. V. Stoecker, G. A. Hance, and S. E. Umbaugh, "Automatic color segmentation algorithms," *IEEE Engineering in Medicine and Biology Magazine*, vol. 9, pp. 75–82, 1993.

[11] P. Schmid, "Segmentation of digitized dermatoscopic images by two-dimensional color clustering," *IEEE Transactions on Medical Imaging*, vol. 18, no. 2, pp. 164–171, 1999.

[12] M. Efimenko, A. Ignatev, and K. Koshechkin, "Review of medical image recognition technologies to detect melanomas using neural networks," *BMC Bioinformatics*, vol. 21, no. S11, p. 270, 2020.

[13] A. G. C. Pacheco and R. A. Krohling, "Recent advances in deep learning applied to skin cancer detection," in *Proceedings of the Neural Information Processing Systems (NeurIPS)*, Retrospectives Workshop, Vancouver, Canada, December 2019.

[14] M. Goyal, T. Knackstedt, S. Yan, and S. Hassanpour, "Artificial intelligence-based image classification for diagnosis of skin cancer: challenges and opportunities," *Computers in Biology and Medicine*, vol. 127, pp. 1–9, 2019.

[15] A. Esteva, B. Kuprel, R. A. Novoa et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.

[16] H. A. Haenssle, C. Fink, R. Schneiderbauer et al., "Man against Machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, vol. 29, no. 8, pp. 1836–1842, 2018.

[17] T. J. Brinker, "ScienceDirect A convolutional neural network trained with dermoscopic images performed on par with 145 dermatologists in a clinical melanoma image classification task," *European Journal of Cancer*, vol. 111, pp. 148–154, 2019.

[18] A. G. C. Pacheco and R. A. Krohling, "The impact of patient clinical information on automated skin cancer detection," *Computers in Biology and Medicine*, vol. 116, Article ID 103545, 2020.

[19] M. A. Kadampur and S. Al Riyaee, "Skin cancer detection: applying a deep learning based model driven architecture in the cloud for classifying dermal cell images," *Informatics in Medicine Unlocked*, vol. 18, Article ID 100282, 2020.

[20] S. Jinnai, N. Yamazaki, Y. Hirano, Y. Sugawara, Y. Ohe, and R. Hamamoto, "The development of a skin cancer classification system for pigmented skin lesions using deep learning," *Biomolecules*, vol. 10, no. 8, pp. 1123–1213, 2020.

[21] L. Wei, K. Ding, and H. Hu, "Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network," *IEEE Access*, vol. 8, pp. 99633–99647, 2020.

[22] C. T. Pham, M. C. Luong, D. Van Hoang, and A. Doucet, "AI outperformed every dermatologist: improved dermoscopic melanoma diagnosis through customizing batch logic and loss function in an optimized Deep CNN architecture," vol. 11, pp. 1–21, 2020, http://arxiv.org/abs/2003.02597.

[23] A. Udrea, G. D. Mitra, D. Costea et al., "Accuracy of a smartphone application for triage of skin lesions based on machine learning algorithms," *Journal of the European Academy of Dermatology and Venereology*, vol. 34, no. 3, pp. 648–655, 2020.

[24] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning," *IEEE Access*, vol. 8, pp. 114822–114832, 2020.

[25] M. A. Anjum, J. Amin, M. Sharif, H. U. Khan, M. S. A. Malik, and S. Kadry, "Deep semantic segmentation and multi-class skin lesion classification based on convolutional neural network," *IEEE Access*, vol. 8, pp. 129668–129678, 2020.

[26] N. Gessert, M. Nielsen, M. Shaikh, R. Werner, and A. Schlaefer, "Skin lesion classification using ensembles of

multi-resolution EfficientNets with meta data," *Methods (Duluth)*, vol. 7, Article ID 100864, 2020.

[27] E. Goceri, "Deep learning based classification of facial dermatological disorders," *Computers in Biology and Medicine*, vol. 128, Article ID 104118, 2021.

[28] S. Karki, P. Kulkarni, and A. Stranieri, "Melanoma classification using EfficientNets and Ensemble of models with different input resolution," in *Proceedings of the ACSW '21: 2021 Australasian Computer Science Week Multiconference*, pp. 1–5, Dunedin New Zealand, February 2021.

[29] A. G. C. Pacheco, G. R. Lima, L. A. Dos Santos de Melo et al., "PAD-UFES-20: a skin lesion dataset composed of patient data and clinical images collected from smartphones," *Data in Brief*, vol. 32, pp. 1–8, 2020.

[30] F. Sultana, A. Sufian, and P. Dutta, "A review of object detection models based on convolutional neural network," in *Intelligent Computing: Image Processing Based Applications*, J. Mandal and S. Banerjee, Eds., Springer, Singapore, pp. 1–16, 2020.

[31] Q. V. Le and M. Tan, *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*, 2019.

[32] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, San Francisco, CA, USA, August 2016.

[33] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le, "Self-training with noisy student improves imagenet classification," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, Seattle, WA, USA, June 2020.

[34] ISIC-2019, "ISIC-2019-Skin lesion analysis towards melanoma detection," 2019, https://challenge2019.isic-archive.com/.