MULTIDISCIPLINARY SCIENCE JOURNAL

RESEARCH ARTICLE
Published Online: January 25, 2024
https://doi.org/10.31893/multiscience.2024117

Enhancing skin cancer classification using a fusion of Densenet and Mobilenet models: a deep learning ensemble approach

Md Hasan Imam[a] | Nazmun Nahar[a] | Md Auhidur Rahman[a] | Fazle Rabbi[a]

[a]Noakhali Science and Technology University, Noakhali, Bangladesh.

Abstract Skin cancer is a condition that causes the formation of abnormal cells in the skin tissue. Every year, millions of individuals experience skin cancer. Skin cancer identification in its early stages is an expensive and difficult process. Currently, deep learning models have revealed promising results in automated skin cancer classification. The goal of this study is to identify benign or malignant skin lesion using the Internation Skin Imaging Collaboration (ISIC) dataset and a concatenated model that combines two powerful Convolutional Neural Network (CNN) architectures, DenseNet and MobileNet. The ISIC dataset, which consists of a large collection of dermoscopic images, provides a valuable resource for training and evaluating skin cancer classification models. By concatenating the feature maps of DenseNet and MobileNet, the proposed model capitalizes on their individual strengths. The concatenated model is trained using a combination of techniques called transfer learning and fine-tuning, which leverage the pre-trained weights of the individual models and adapt them to the skin cancer classification task. On the test dataset, our proposed model finally achieves 93.75% accuracy, which is 4.69% higher than DenseNet, 3.13% higher than MobileNet, and 3.13% higher than the combination of ResNet50 and InceptionNetV3. The execution time of our proposed ensemble method is 49.48 min, and the execution times of MobileNet and DenseNet are 178.9 and 170.5 min, which are more than our proposed ensemble approach.

Keywords: image processing, CNN, melanoma of the skin

1. Introduction

Due to significant advances in deep learning and computer vision in recent years, the medical diagnostic industry has undergone a revolution. Skin cancer, one of the most prevalent cancers in the world, has a growing incidence rate and serious public health concerns (Craythorne et al., 2017). The Malignant melanoma incidence is increasing at a faster rate than other cancers (Gloster et al., 2006). By January 12, 2023, the Cancer Society of America projects that there will be more than 97,610 new cases of melanoma, with an estimated 7,990 fatalities (American Cancer Society, 2023). Melanoma incidence has risen dramatically in recent decades (American Cancer Society, 2023). To improve patient outcomes and lower death rates, it is essential to identify skin lesions as early as possible and to classify them accurately. In particular, deep learning has become a potent technique for accurately identifying different types of skin cancer by extracting complex patterns and features from medical images.

A crucial component of deep learning models' success is their capacity to learn hierarchical representations directly from unprocessed data, eliminating the requirement for feature engineering by hand. Convolutional Neural Networks (CNNs), a subset of deep learning models created especially for image analysis, have proven to perform incredibly well in picture classification tasks, including the classification of skin cancer (Sultana et al., 2018). Because of their particular qualities that improve feature extraction and computational efficiency, DenseNet and MobileNet have emerged as leading candidates among the many CNN architectures.

In this context, this study proposes an effective method for improving the accuracy and robustness of skin cancer classification by fusing the DenseNet and MobileNet models. We use a deep learning ensemble technique that takes advantage of the advantages of both architectures' complementary feature extraction abilities. Individual learners' performance is limited to decision-making on sensitive problems. This problem can be handled by merging the decisions of individual learners. The combined decision is predicted to be more accurate than that of individual learners. We aimed to enhance the discriminative power and robustness of the skin cancer classification system by combining these two models.

In this study, we performed experiments using the largest collection of dermoscopic photos of skin lesions that have been labeled as benign or malignant (ISIC2017) from the Kaggle competition (Skin Cancer: Malignant vs. Benign, 2023). We seek to determine the effectiveness of the ensemble strategy in the context of actual clinical settings by using the ISIC2017 dataset.

The following contribution is made by this work to the detection of skin cancer.

-Multiple pretrained models, including MobileNet and DenseNet, are used in the proposed approach for feature extraction. The distinctive features of the MobileNet and DenseNet models are combined.

-The combined features are sent to multiple completely connected nonlinear layers. The features were then input into a classifier for the classification of skin diseases.

-To demonstrate the generalizability of the proposed approach, an in-depth examination and analysis of the deep learning-based pretrained models was performed on the skin cancer dataset.

-The results of the proposed approach for detecting skin cancer are compared to other individual learners and the combination of ResNet50 and InceptionV3, comparable methods demonstrate that the proposed method is reliable and can deliver improved results.

-We also compared the execution time of our proposed model with individual learners and the combination of ResNet50 and InceptionV3.

The remaining parts are categorized into the following groups: Section 2 provides a comprehensive analysis of the existing literature on skin cancer. A comprehensive analysis of the datasets and the proposed methodology is presented in Section 3. Section 4 provides detailed information regarding statistical measurements. Furthermore, Section 4 contains the analysis of the experimental data, the discussion of the research findings, and the subsequent issues involving them. Section 5 includes a comprehensive summary and future work.

2. Literature Review

A tremendous amount of work has been done to identify skin cancer using machine learning as well as deep learning methodologies.

The ISIC2020 dataset is trained and evaluated using a DenseNet model by authors (Zhang et al., 2021). In addition, the writer contrasted the performance of this model with that of the VGG and ResNet method. In comparison to VGG and ResNet models, their accuracy rate utilizing DenseNet is 92.5%, which is a better performance. Developed an image processing, deep learning, and transfer learning strategy (Rezaoana et al., 2020) for the ISIC dataset, which contains 25,780 pictures of benign and malignant. Their accuracy rate in this approach is 79.45%, which is higher than VGG-16 and VGG-19. With a dataset of 600 images, authors suggested (Hartanto et al., 2020) MobileNetV2 and Faster R-CNN algorithms that are used and implemented on an Android-based application that can identify actinic keratosis or melanoma, two types of skin cancer. Their research indicates that Faster R-CNN outperformed MobileNetV2 by 87.2% in terms of accuracy. Dermatofibroma (DF), keratosis-like lesions (BKL), and basal cell carcinoma (BCC) are three separate kinds of skin cancer that have been found, and (Zhang et al., 2020) suggested a method for classifying them using the HAM1000 algorithm. The scientists classified three skin cancer lesions with an accuracy of 98.5% without shaving the lesions using pre-trained VGG19 and a softmax layer in place of the final deep CNN layer. The International Skin Imaging Collaboration (ISIC) dataset was used by authors (Aima et al., 2019) to propose a method for delineating the border of the lesion in dermoscopic pictures using CNN with stochastic gradient descent. This process achieved a classification accuracy of 74.76% when tested on a dataset of 514 dermoscopic images. High computational speed in training data and increasing validation loss is the suggested technique. The proposed model (Janoria et al. 2020) has been performed on ISIS datasets. According to the experimental findings, the VGG 16 CNN model combined with the K-Nearest Neighbour algorithm produced the maximum accuracy of 99%.The performance of sophisticated models, such as ensemble learning utilising boosted trees, is less than 50%, which is not up to par. Therefore, the aforementioned dataset is highly suited for linear binary classifiers but not for coarse multiclass models. For accurate lesion border segmentation, researchers used Mask R-CNN with DeeplabV3+ on ISIC-2017 dataset in their study (Goyal et al., 2020), and achieved high sensitivity (89.93%) and specificity (97.94%). Their Ensemble-A technique performed more sensitively than FrCN, FCNs, U-Net, and SegNet by 4.4%, 8.8%, 22.7%, and 9.8%, respectively. On the same dataset, Ensemble-S outperformed other architectures in terms of specificity (97.98% for benign cases, 97.30% for melanoma, and 98.58% for seborrheic keratosis). Three layers that are concealed with output widths of 16, 32, and 64 make up the suggested model (Fu'adah et al., 2020). Several optimizers, including SGD, RMSprop, Adam, and Nadam, were tested using a learning rate of 0.001. When classifying skin lesions into four groups (dermatofibroma, nevus pigmentosus, squamous cell carcinoma, and melanoma), Adam optimizer had the maximum accuracy of 99%. These outcomes fared better than the method used to classify skin cancer at the moment. Two datasets—ISIC and CPTAC-CM—were used to train the CNNs in the four-layer CAD system that was proposed (Diab et al., 2022) where investigated and contrasted were GoogleNet, ResNet-50, AlexNet, and VGG19. The suggested CAD system has a 99.8% accuracy rate for the ISIC database and a 99.9% accuracy rate for the CPTAC-CM database. The multi-classification and grading of various forms of skin cancer will be the main emphasis of the effort. Reducing computing time and utilising diverse datasets with more image data present further challenges. One non-cancer type was picked from Human against Machine (HAM10000) for classification in this work (Abuared et al., 2020) which employs VGG19 and transfer learning techniques. The model assessed the overall accuracy and loss of the network. A lesion classification network and a feature discrimination network are used in the proposed model (Wei et al., 2020) to construct an efficient semantic segmentation technique for lesion area segmentation in dermoscopy pictures. This strategy outperforms deep learning algorithms in melanoma diagnosis without complicated image processing using the ISBI 2016 skin lesion study dataset. The authors (Daghrir et al., 2020) describe a hybrid approach for detecting melanoma skin cancer that can be applied to any suspicious site. Essentially, that's a projection based on three distinct methods: A convolutional neural network and two traditional machine learning classifiers were trained using a set of features that describe the borders, texture, and colour of a skin lesion. Integrating all of these strategies increases accuracy. This research (Shorfuzzaman et al., 2022) provides an explainable CNN-based stacked ensemble architecture for early detection of melanoma skin cancer. The transfer learning principle is applied in this architecture, where numerous CNN sub-models that execute the same classification job are built. A new model known as a meta-learner takes all of the predictions from the sub-models and produces the final forecast of the prediction. The ensemble model performs well in terms of accuracy (95.76%), sensitivity (96.67%), and AUC (0.957) with the evaluation findings.

These days, researchers frequently employ ensemble networks to boost classification performance. Typically, each model is trained separately, and predictions from different models are combined to provide the results. The authors propose (Rathore et al., 2023) a CNN-FFNN model that combines CNN and FFNN to predict knee osteoarthritis. The output of each of the four deep neural network architectures utilised by the author (Harangi, 2018) (GoogleLeNet, AlexNet, ResNet, and VGGNet) is combined as part of an ensemble technique. In practice, they advocate creating a framework from a number of strong convolutional neural networks (CNNs), with the framework's weighted output being used to perform the final classification. On the ISBI datasets, this model has 89% accuracy, which is superior to individual networks' lower accuracy. Suggested (Imran et al., 2022) a VGGNet, CapsNet, and ResNet ensemble for skin cancer diagnosis utilizing the ISIC public dataset. The suggested model achieves 93.5% accuracy, while VGGNet, CapsNet, and ResNet achieve 79%, 75%, and 69% accuracy, respectively. An ensemble technique is proposed (Alizadeh et al., 2021) in which two CNNs are utilised to categories images in the CNN phase and image texture feature extraction is employed to improve classification performance. The results of each phase are then integrated to arrive at the final diagnosis. The results reveal that the ensemble method outperforms each phase individually in terms of evaluation measures. The drawbacks of this method do not use extra features like color and shape characteristics for detecting melanoma. An approach is proposed (Pacheco et al., 2021) in which metadata are employed to aid data categorization by improving the most relevant characteristics retrieved from images along the classification process. In this work, a combined technique based on MetaNet and one based on feature concatenation achieve 90.9% accuracy.The suggested ensemble technique (Codella et al., 2017) seeks to improve melanoma detection by improving skin lesion segmentation while analysing both the detected lesion region and the nearby tissue. The methodology employs a hand-coded feature extractor, sparse-coding techniques, deep residual networks, fully convolutional neural networks, and Support Vector Machines (SVM). These several components work together to provide robust melanoma detection and segmentation within dermoscopy pictures.

3. Methodology

3.1. Dataset

The dataset, which primarily consists of the 'International Skin Imaging Collaboration (ISIC)' archive data, is downloaded from Kaggle and consists of benign and malignant images. The ISIC archive contains dermoscopic images acquired using a specialized medical imaging technique known as dermoscopy. Dermoscopy images (Celebi et al., 2019) are used to diagnose and identify skin cancer and other dermatological disorders in their early stages. Dermoscopic lesion images in JPG format comprise the dataset. ISIC is an academic-industry collaboration that accelerate the development of digital skin imaging solutions that can shorten the melanoma life cycle. The dataset from the ISIC repository is commonly used in academic research, especially in the fields of computer vision and machine learning. To aid in the early detection and diagnosis of skin cancer, researchers are using this dataset to develop and test models and algorithms that automatically analyze dermoscopic images (Wen et al., 2022).

There are 3297 total images in the dataset. There are 1440 benign and 1197 malignant photographs in the train folder, whereas there are 360 benign and 300 malignant images in the test folder. The training and test datasets are visually depicted in Figure 1 and 2, respectively. The benign and malignant dermoscopy image samples are presented in Figure 3 and 4, respectively. Here we used only two types of skin cancer: benign and malignant. Because the ISIC dataset contains almost equal numbers of benign and malignant classes, there is no bias in the dataset.

3.2. Data Pre-processing

The ISIC Challenge dataset comprises images of dermoscopic skin lesions acquired using different dermoscopic and camera devices from around the world. Therefore, it is crucial to preprocess these images. A variety of preprocessing processes are conducted to improve image quality and highlight important characteristics in this study. Resizing the photos is the initial step after gathering the skin cancer images from various sources and sizes. Shape conversion: (224, 224, 3). In the second stage, we removed the hair from the dermoscopy images. In this study, hairs in dermoscopy images are eliminated using an algorithm called DullRazor (Lee et al., 1997). Figure 5 shows the DullRazor algorithm procedure for eliminating hair from the original image.

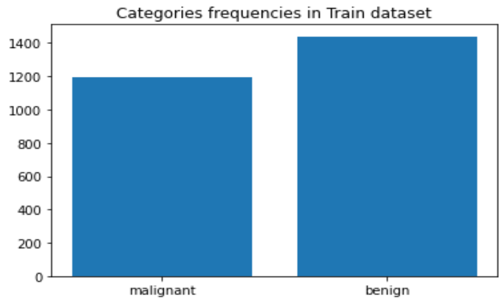

Figure 1 Train dataset.

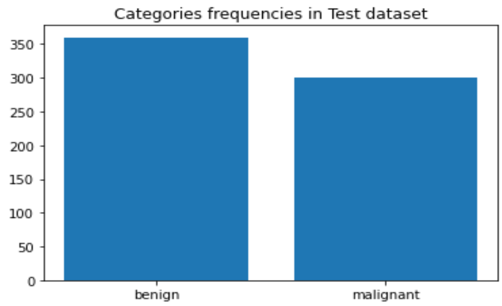

Figure 2 Test dataset.

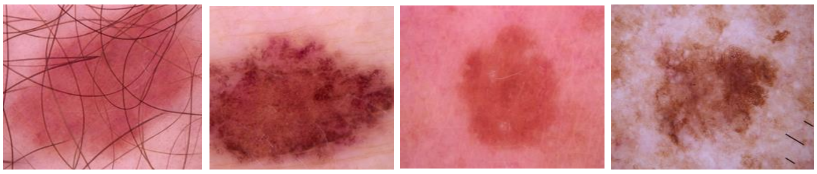

Figure 3 The dermoscopic image sample of Benign.

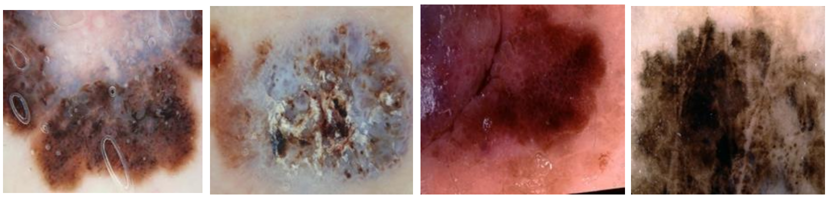

Figure 4 The dermoscopic image sample of Malignant.

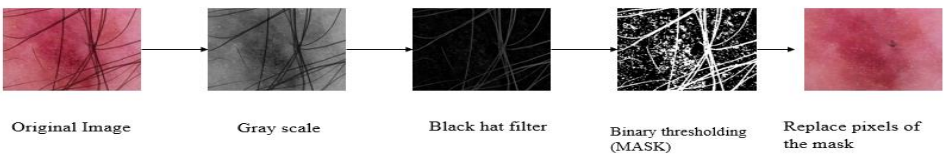

Figure 5 Hair removal using DullRazor algorithm: before and after.

3.3. Optimal Hyperparameter of our Proposed Ensemble Model

The loss function that we employ to train our model is called Binary Cross-Entropy. Because there are no longer significant differences in the degrees of accuracy achieved during training and validation, the model we use has been trained for a maximum of fifty epochs. The optimizer developed by Stochastic Gradient Descent (SGD) (Ketkar et al., 2017) is used to optimize the loss function. Table 1 illustrates our proposed method for optimally adjusted hyperparameters. In this test, there are a total of fifty epochs, and each batch has a size of thirty-two. We also evaluate our model using other hyperparameters. We trained our model using the Adam SGD and RMSProp optimizer. Among them, the SGD optimizer provides the optimal output. We also train our model with different batch sizes and different numbers of epochs. Among them, we found optimal batch sizes of 32 and 50 epochs. We used the same test data for the validation process and the same hyperparameter for the validation technique.

Table 1 Value of hyperparameters.

Hyperparameter Name	Value
Batch Size	32
Epochs	50
Loss Function	Binary Crossentropy
Optimizer	Stochastic Gradient Descent (SGD)
Learning rate	0.001

3.4. Proposed Method

In this article, we suggest an ensemble technique to categorize skin cancer, as shown graphically in Figure 6. Our suggested approach is explained in the following subsections.

3.5. Selection Base Learners

Although the choice of the base learner varies depending on the situation, the key is to select the models that are most suited to the situation. As base learners, we chose two distinct methods for categorization, both of which are state-of-the-art CNN frameworks. MobileNet (Michele et al., 2019) and Densenet121 (Zhu et al., 2017) are these base learners.

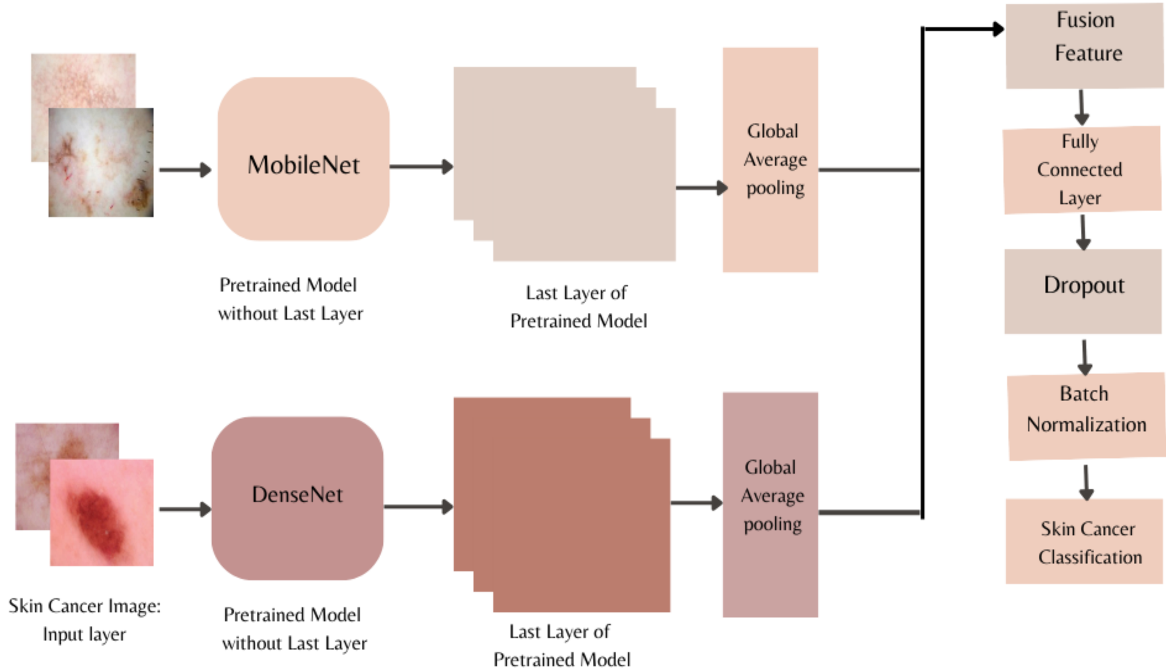

Figure 6 Ensemble deep learning approach for skin cancer detection.

3.6. Training the Base Learner Model and Calculation of their Weights

After each model is chosen, it is trained separately on the dataset, and its percentage accuracy (*acc_j*) is measured on the validation set. The validation dataset was not used during the training process. The same probability is learned separately by each model because each model uses a distinct set of hyper-parameters. Models can continue to be trained until they reach convergence or until there is no reduction in the loss value. After these models have been trained, they are then assessed using a validation dataset, and the accuracy levels measured. The weight ratio "*j*" is computed for each individual model using the estimated accuracies in Equations (1) and (2). In this case, *j*' ranges from *1 to m*, with m representing the total number of models.

$$\varphi_i = acc_j - \beta + 1 \qquad (1)$$

$$\beta = min[acc_j]_{j=1}^{m} \qquad (2)$$

3.7. Aggregation of Ensemble Model

The trained models are now used to generalize or estimate the input data. The result of the models is taken after applying the 'Sigmoid'' function. The model output for all input data T_i represents the probability vectors C_k. Where 'k' is taken from 1...nc and 'nc' stands for all classes. The model's output has been multiplied by the appropriate multiplication factor j. As demonstrated in Eq. (3), the calculated probabilities for each class are added collectively to form the result for the ensemble approach for the class. Algorithm 1 depicts the procedure for the weighted ensemble framework, making it simple to learn and apply. Equation (3) is used to calculate the final result for a particular class.

$$Output_{E_i,ensemble} = max\left[\sum_{i=1}^{m} \varphi_i \times [O_k]_{j,i}\right]_k^c \quad (3)$$

The proposed ensemble model is described in Algorithm 1. Algorithm 1 describes how we are going to split our dataset into training and validation. The first part is used as training data, and is loaded into S = (x_1, y_1),..., (x_n, y_n), where x_n is the set of feature vectors and y_n is the label. The validation dataset is then entered into V, and a subset of the dataset is used. Each of the m models is individually trained using dataset S. Here, m equals to two. Each model was then used to make predictions on $[V_i]_g$ i=1 when training was complete. To determine the accuracy of the model M_j , we use these predictions.

Algorithm 1: Ensemble Model for Skin Cancer	
Input	*Skin cancer datasets S={(X_1,Y_1),(X_2,Y_2).......(X_n , Y_n)}* *V={(X_1,Y_1),(X_2,Y_2).......(X_g,Y_g)}for training and testing respectively.*
Output	*Ensemble model M.*
Process	*1. for do j=1 to m* *a) Train model MobileNet and DenseNet using dataset S.* *b) Calculate the class probability for p_j using the trained model of MobileNet and DenseNet.* *c) Calculate output class of MobileNet and DenseNet* *d) Calculate accuracy acc_j of model M_j on V* *2. for do j=1 to m* *a) Calculate weight of model M_j* *3. Compute the output of the proposed model.* *4. Calculate the accuracy of the proposed ensemble method .*

4. Result and Discussion

In this section, the experiments that we conducted and the results of those experiments are described.

4.1. System Configuration

To execute our proposed ensemble method and transfer learning in this problem, we employed a tensor-flow neural network (Pang et al., 2020). The primary reason for employing this network is the presence of several matrix multiplications. We encountered some challenges during our work, which were primarily caused by the limitations of CPU processing. As a solution, we use Google's collaborative cloud server. After that, both the central processing unit (CPU) and the Jupyter Notebook can be operated easily. After employing these techniques, we can proceed to train and assess the effectiveness of our suggested deep learning methodology.

4.2. Various Performance Evaluation Metrics

Several parameters for performance are described in the standard evaluation of the classification model. Classification accuracy is the most frequently used statistic. A high degree of accuracy indicates that the classifier correctly identified all normal skin samples as normal and all cancerous skin samples as cancerous. Calculating accuracy, recall (or sensitivity), and precision, which are all crucial metrics to consider while working with classification issues, may be done using the following equations. The total number of true positives, false positives, true negatives, and false negatives that were identified is represented by the abbreviations TP, FP, TN, and FN, respectively. The F-score, which is a helpful statistical tool for classification, is derived by taking the harmonic mean of the accuracy and recall values. The following equation may be used to determine the levels of accuracy, precision, and recall, as well as F-score.

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

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

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

$$F - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

4.5. Experimental Result

We showed a study of how well the proposed ensemble method, which combines MobileNet and DenseNet, works. The ISIC public dataset for classifying skin cancer from dermoscopic images is used to test the effectiveness of the model. We test the training with different batch sizes and choose the best hyperparameters for the suggested model, which are 40 epochs, an average learning rate of 0.001 with an average batch size of 32, and a stochastic gradient descent (SGD) optimizer.

Table 2 shows the training and testing accuracy of the proposed model using different optimizers. From the table, Adam optimizer training and testing accuracy are 98.44% and 85.45%, respectively, and the training and testing losses are 0.031 and 0.450, respectively. The RMSprop optimizer has the same training accuracy but the testing accuracy is different from that of the Adam optimizer. The testing accuracy of the RMSprop optimizer is 88.67%. Training and testing losses are 0.078 and 0.277, respectively, which are less than those of the Adam optimizer. The SGD optimizer training and testing accuracy are 95.31% and 93.75%, respectively. The training and testing losses of the SGD optimizer are 0.131 and 0.186, respectively.
Here, we can say that the overall SGD optimizer performs better than the other two optimizers.

Table 2 The training and testing accuracy of the model using several optimizers.

Optimizer	Accuracy of Training	Accuracy of Testing	Loss of Training	Loss of Testing
Adam	98.44%	85.45%	0.031	0.450
RMSProp	98.44%	88.67%	0.078	0.277
SGD	95.31%	93.75%	0.131	0.186

Table 3 shows the training and testing accuracy and the loss over the epoch. From the table, we can see that when the epoch number is increased, the training and testing accuracy also increases and training and testing loss decreases. At 1 epoch, the training and testing accuracy was 79.4% and 81.9% respectively. In the same epoch the training and testing losses were 0.443 and 0.372, respectively.

When the epoch number increases, accuracy also increases and loss decreases. At epoch number 30, the training accuracy is 90.4% and the testing accuracy is 85.3%. The training loss is 0.226 and the testing loss is 0.330.

Table 3 Performance Comparison of Proposed Ensemble Method over Epoch.

Epoch	Accuracy of Training	Accuracy of Testing	Loss of Training	Loss of Testing
1	0.794	0.819	0.443	0.372
10	0.869	0.873	0.294	0.307
20	0.892	0.869	0.239	0.282
30	0.904	0.853	0.226	0.330
40	0.9071	0.8727	0.2109	0.2746
50	0.9120	0.8660	0.1971	0.2784

Table 4 presents a comparative analysis of the performance of the proposed and other classifier models. We compared our proposed model with two individual learners and a combined model (ResNet50 (Mascarenhas et al., 2021) and InceptionV3 (Szegedy et al., 2016)). We evaluated the proposed model using several different measures, including accuracy, loss, precision, recall, F1-Score, and area under the curve (AUC). The results for MobileNet's accuracy, recall, F1-score, and AUC are as follows: 90.4%, 0.287, 0.937, and 0.947, respectively. The loss for DenseNet is 0.272, and its accuracy is 98.06%. DenseNet has 0.828 precision, 0.966 recall, 0.638 F1-score, and an AUC of 0.965. The combined accuracy and loss of the ResNet50+InceptionV3 model are 90.62% and 0.265, respectively. The accuracy and loss of our proposed ensemble (MobileNet+DenseNet) method are 93.75 and 0.186, respectively. Our proposed ensemble method achieves 0.967 precision, 0.909 recall, 0.680 F1-score, and 0.986 area under the curve (AUC). Our proposed ensemble technique outperforms the other methods.

Table 4 Evaluating the Proposed Ensemble Method's Performance Compared to Other Classifiers.

Model	Accuracy	Loss	Precision	Recall	F1-Score	AUC
MobileNet	90.62%	0.287	0.937	0.882	0.669	0.947
DenseNet	89.06%	0.272	0.828	0.966	0.638	0.965
Deep Ensemble Approach	90.62%	0.265	0.992	0.778	0.593	0.972
Proposed Ensemble Method (MobileNet+DenseNet)	93.75%	0.186	0.967	0.909	0.680	0.986

Figure 7 shows the accuracy curve of the MobileNet Classifier. At epoch 1, the initial accuracy of the MobileNet classifiers was approximately 72% and 80% for training and validation, respectively. After increasing the number of epochs, the classifier accuracy is also increased for both training and validation. Figure 7 also shows the loss curve of that classifier. At epoch number 1, the losses are approximately 0.55 and 0.43 for training and validation, respectively. When the number of epochs is increased, the loss decreases for both training and validation. At epoch number 50, the losses are 0.30 and 0.26 for both training and validation.

The accuracy curve of the DenseNet Classifier is depicted in Figure 8. The initial accuracy of the DenseNet classifier is approximately 73% for training and 84% for validation at the beginning of epoch 1. When the number of epochs increases, the accuracy of the classifier also increases, and this holds true for both training and validation. The loss curve for the classifier is shown in Figure 8. At the first epoch, the losses are around 0.55 for training and 0.39 for validation. The loss is reduced not just during training but also during validation when the number of epochs used is increased. When we reach epoch number 50, the losses are approximately 0.30 and 0.23 for the training and validation sets, respectively.

Figure 7 Accuracy and Loss of MobileNet.

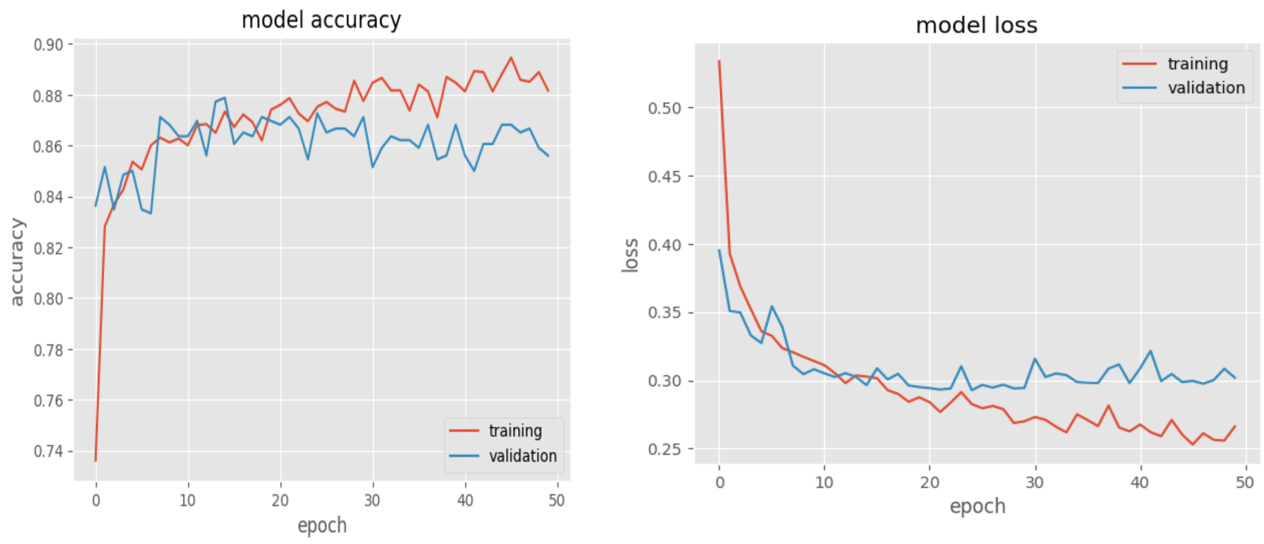

Figure 8 Accuracy and Loss of DenseNet.

The accuracy curve of the integrated ensemble classifiers (ResNet50+InceptionV3) is shown in Figure 9. The initial accuracy of combined classifiers is roughly 66% for training and 68% for validation at the beginning of epoch 1. When the number of epochs used in the training and validation processes is raised, the classifier's accuracy likewise rises to a higher level. The loss curve of the classifier is also depicted in Figure 9. In the first epoch, the losses are around 0.81 and 0.69, respectively, for training and validation. Both the training and validation losses can be reduced by increasing the number of epochs used in the process. The losses for the training and validation sets are 0.27 and 0.49, respectively, at epoch number 50.

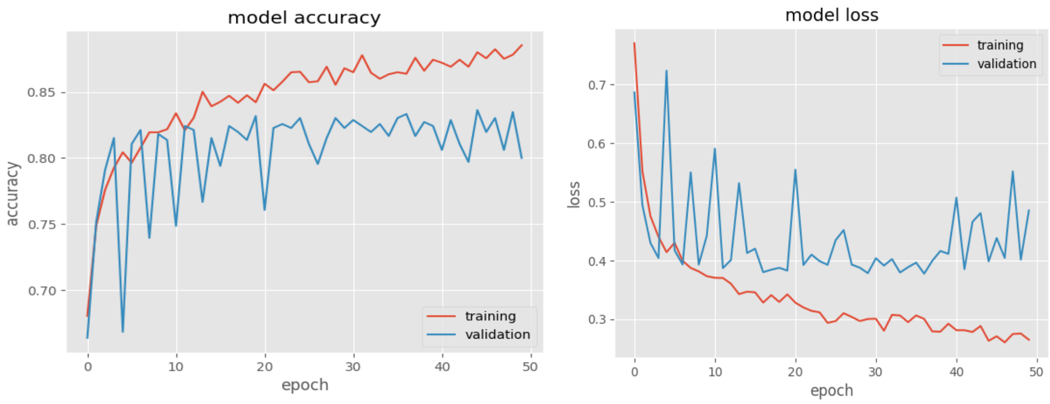

Figure 9 Accuracy and Loss of Combined Model (ResNet50+InceptionV3).

Figure 10 illustrates the proposed method's accuracy curve. MobileNet's training and validation accuracy are approximately 78% and 82% at epoch 1, respectively. The training and validation accuracy of the classifier improved with more epochs. The gap between training and validation accuracy widens when more epochs are used. The loss curve for that classifier is shown as well in Figure 10. The losses in training are about 0.38 and in validation they are about 0.44 at epoch 1. The loss during training and validation is reduced as the number of epochs increases. At epoch 50, the losses in training and validation are 0.19 and 0.26, respectively. The gap between training and validation loss widens as the epoch increases.

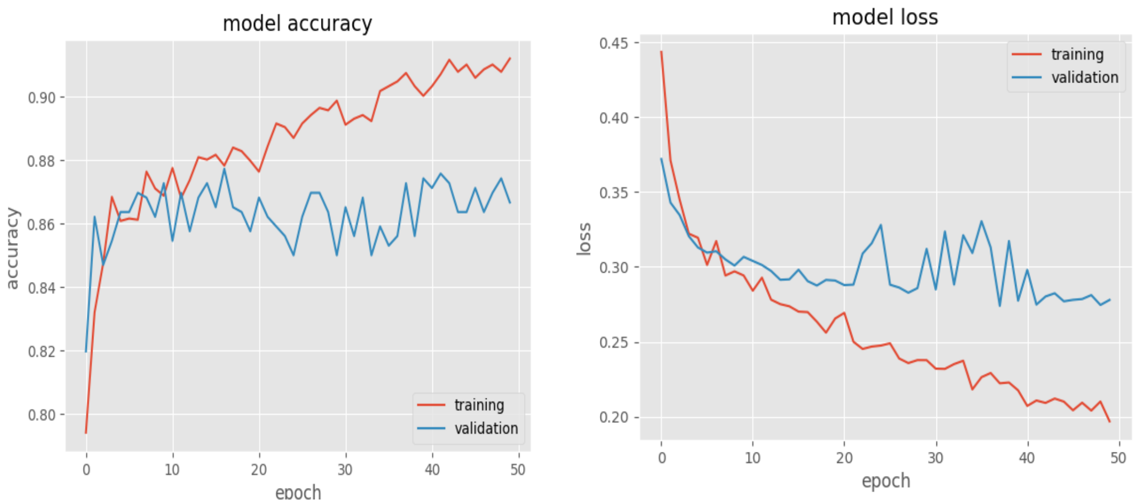

Figure 10 Accuracy and Loss of our Proposed Ensemble Approach.

Figure 11 shows an accuracy comparison of our proposed model with individual learners. From the figure, we can see that the individual learner DenseNet, MobileNet accuracy is lesser than that of our proposed model.

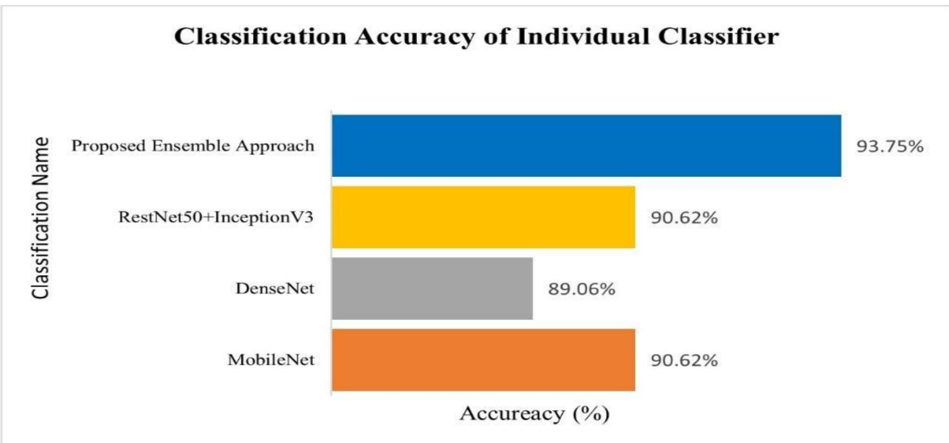

Figure 11 Accuracy Comparison of Our Proposed Model with Individual Classifier.

Table 5 compares the proposed model with certain previously used approaches. On the basis of accuracy, we make comparisons. Our proposed ensemble model (DenseNet + MobileNet) outperforms the state-of-the-art technique, as shown in Table 5.

Table 5 Comparing Our Proposed Model's Efficiency to Current State-of-the-Art Methods.

Reference	Method	Accuracy
(Harangi, 2018)	GoogLeNet+AlexNet+ResNet+VGGNet	89%
(Imran et al., 2022)	VGGNet+CapsNet+ResNet	93.5%
(Alizadeh et al., 2021)	VGG19+ANN	85.2%
(Pacheco et al., 2021)	CNN+MetaNet	90.9%
(Codella et al. 2017)	ResNet+Unet	76%
Proposed Method	DenseNet+MobileNet	93.75%

The total amount of time required to train DenseNet, MobileNet, ResNet50+InceptionV3, and the proposed approach for classification is 178.9, 170.5, 55.3, and 49.48 min, respectively. A graphical representation of the various models is shown in Figure 12. In terms of accuracy and the amount of time required for training, there is a little tradeoff between the suggested model and the ResNet50+InceptionV3 model; however, the proposed model is more heavily weighted due to the greater difference in performance.

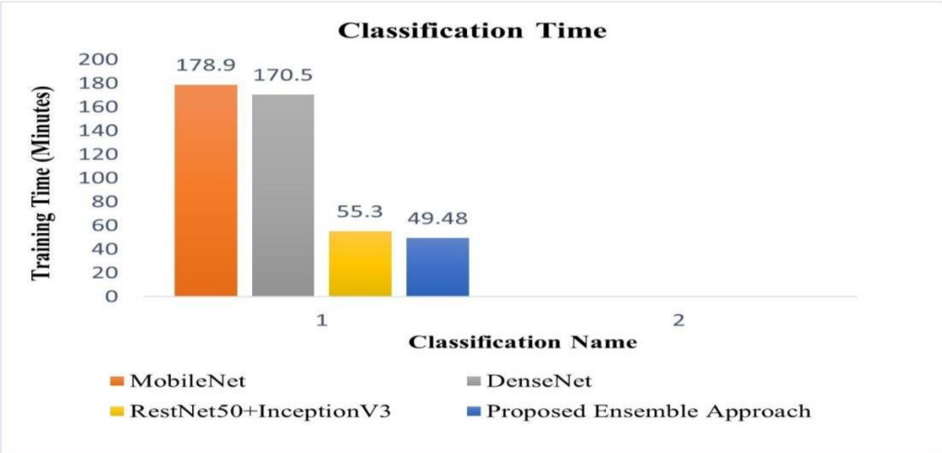

Figure 12 Comparison of our Proposed Model Execution Time with Other Model.

4.6. Discussion

Here, we provide an evaluation of the performance of the proposed combination of MobileNet and DenseNet in an ensemble network. The evaluation of the model's efficacy was performed using the ISIC public dataset, which consists of dermoscopic images used for skin cancer categorization. The training assessment was conducted by adjusting the batch sizes and selecting the optimal hyperparameters, which included 50 epochs, a learning rate of 0.001 with a total number of batches of 32, and a stochastic gradient descent (SGD) optimizer for the proposed model. Figure 11 illustrates the accuracy of the proposed ensemble network compared with that of individual learners. Based on the analysis shown in Figure 11, it can be determined that the proposed model attained the highest accuracy of 93.75%. Figures 7 and 8 display the accuracy and loss curves for the training and testing accuracy of each learner. The proposed ensemble also achieved the best loss function of 0.186. MobileNet, DenseNet, and our proposed ensemble learning models had classification training times of 178.9, 170.5, and 49.48 min, respectively. The training time for our proposed model classification is only 53s per epoch, which is faster than that of any other model. Imran et al. (2022) proposed an ensemble model with a classification training time of 106 s. A graphical representation of several models is also displayed. The proposed approach is significant because of its better performance.

Table 5 compares our proposed methodology with those of other existing studies. From the table, we can see that our proposed ensemble approach achieved the highest accuracy compared with the other ensemble methods because our proposed ensemble method uses Mobilenet and Densenet architectures. MobileNet employs depth-wise separable convolutions, resulting in improved computing efficiency and potentially accelerated inference and training times. The dense connection patterns of DenseNet enable the reuse of features, which can efficiently capture more intricate patterns.

4.7. System Architecture

The overall structure of our proposed model is shown in Figure 13. In this system, the model is loaded when the server starts. Then, the server starts listening for a client request. The flask server functions as a client of the model server whenever a user uploads an image into a web application. The model server invokes the model via a request handler for participation. The predictions are then made by the model and sent to the model server.

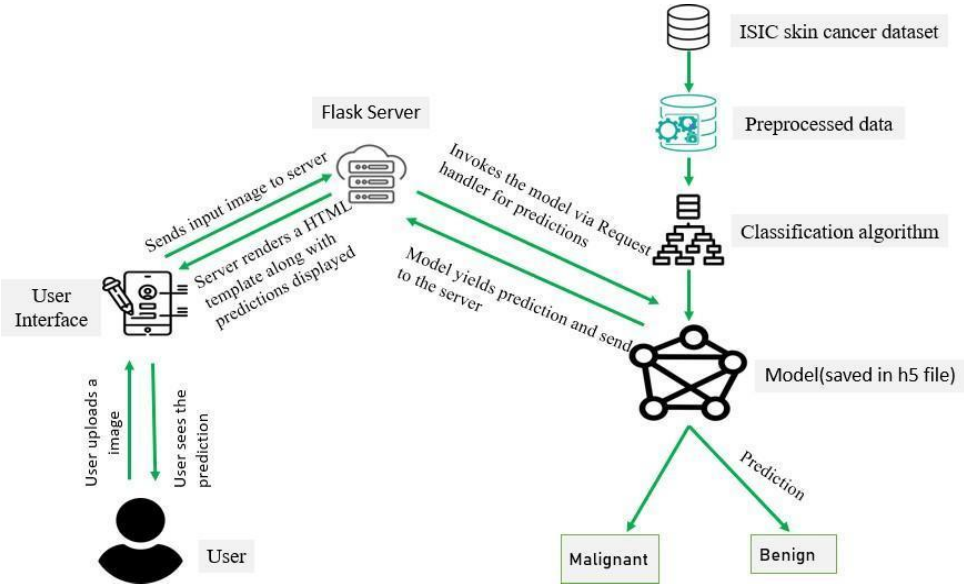

Figure 13 System Architecture.

The following figure shows the result of an image predicted as Benign which is generated by our proposed model.

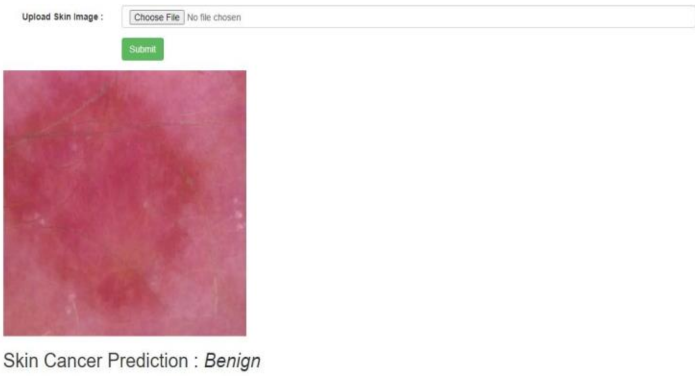

Figure 14 Skin Cancer Prediction.

5. Conclusions

This study proposes a novel approach for skin cancer classification by combining the DenseNet and MobileNet models through concatenation. This approach improves the overall classification results by leveraging the strengths of both models. By addressing the challenges associated with image quality, noise, and artifacts, this study enhances the overall quality of input images for more accurate classification results. This contribution is crucial because high-quality image preprocessing plays a vital role in achieving reliable and robust classification outcomes. It is noticeable from the findings that the suggested ensemble was successful in achieving an average accuracy of 93.75 percent with an average training period of 49.6 minutes. The suggested model achieves higher levels of performance than individual learners in terms of a variety of quality criteria, including sensitivity, accuracy, false-positive rate, specificity, and precision. The integration of Convolutional Neural Networks (CNN) with federated learning can be explored to address data privacy concerns. Federated learning allows training models on distributed data without the need for data sharing, thus preserving privacy. By implementing a federated learning approach, the research can leverage a larger and more diverse dataset from multiple healthcare institutions while ensuring data confidentiality. The proposed method can explore multi-class classification for skin cancer analysis. Although this study focuses on binary classification (malignant vs. benign), expanding it to a multi-class classification approach can enable the identification of different subtypes of skin cancer. This would provide more detailed and comprehensive information for accurate diagnosis and treatment recommendations.

Ethical considerations

To begin with, the ISIC skin cancer dataset is open source and acceptable for research. Second, the ISIC body is very concerned about their data, and they have a privacy working group that is developing privacy guidelines to help assure the confidentiality and appropriate consent for the use and sharing of images and their associated metadata in dermatology imaging. Dermatology imaging presents specific privacy challenges that are not addressed by existing health privacy standards. Total body photography, in particular, makes the individual clearly identifiable due to the presence of their face, characteristic skin blemishes, or tattoos.

Conflict of Interest

The authors declare no conflicts of interest.

Funding

This research did not receive any financial support.

References

Craythorne, E., & Al-Niami, F. (2017). Skin cancer. *Medicine*, *45*(7), 431-434.

Gloster Jr, H. M., & Neal, K. (2006). Skin cancer in skin of color. *Journal of the American Academy of Dermatology*, *55*(5), 741-760.

Melanoma Skin Cancer Statistics. Available in: https://www.cancer.org/cancer/types/melanoma-skin-cancer/about/key-statistics.html Accessed on: August 18,2023.

Sultana, F., Sufian, A., & Dutta, P. (2018, November). Advancements in image classification using convolutional neural network. In *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)* (pp. 122-129). IEEE.

Skin Cancer: Malignant vs. Benign. Available in: https://www.kaggle.com/datasets/fanconic/skin-cancer-malignant-vs-benign. Accessed on: June 17,2023

Zhang, Y., & Wang, C. (2021, March). SIIM-ISIC melanoma classification with DenseNet. In *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)* (pp. 14-17). IEEE.

Rezaoana, N., Hossain, M. S., & Andersson, K. (2020, December). Detection and classification of skin cancer by using a parallel CNN model. In *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)* (pp. 380-386). IEEE.

Hartanto, C. A., & Wibowo, A. (2020, September). Development of mobile skin cancer detection using faster R-CNN and MobileNet v2 model. In *2020 7th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)* (pp. 58-63). IEEE.

Zhang, N., Cai, Y. X., Wang, Y. Y., Tian, Y. T., Wang, X. L., & Badami, B. (2020). Skin cancer diagnosis based on optimized convolutional neural network. *Artificial intelligence in medicine*, *102*, 101756.

Aima, A., & Sharma, A. K. (2019, February). Predictive approach for melanoma skin Cancer detection using CNN. In *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur-India*.

Janoria, H., Minj, J., & Patre, P. (2020, November). Classification of skin disease from skin images using transfer learning technique. In *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 888-895). IEEE.

Goyal, M., Oakley, A., Bansal, P., Dancey, D., & Yap, M. H. (2019). Skin lesion segmentation in dermoscopic images with ensemble deep learning methods. *IEEE Access*, *8*, 4171-4181.

Fu'adah, Y. N., Pratiwi, N. C., Pramudito, M. A., & Ibrahim, N. (2020, December). Convolutional neural network (cnn) for automatic skin cancer classification system. In *IOP conference series: materials science and engineering* (Vol. 982, No. 1, p. 012005). IOP Publishing.

Diab, A. G., Fayez, N., & El-Seddek, M. M. (2022). Accurate skin cancer diagnosis based on convolutional neural networks. *Indonesian Journal of Electrical Engineering and Computer Science*, *25*(3), 1429-1441.

Abuared, N., Panthakkan, A., Al-Saad, M., Amin, S. A., & Mansoor, W. (2020, November). Skin cancer classification model based on VGG 19 and transfer learning. In *2020 3rd International Conference on Signal Processing and Information Security (ICSPIS)* (pp. 1-4). IEEE.

Wei, L., Ding, K., & Hu, H. (2020). Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network. *IEEE Access*, *8*, 99633-99647.

Daghrir, J., Tlig, L., Bouchouicha, M., & Sayadi, M. (2020, September). Melanoma skin cancer detection using deep learning and classical machine learning techniques: A hybrid approach. In *2020 5th international conference on advanced technologies for signal and image processing (ATSIP)* (pp. 1-5). IEEE.

Shorfuzzaman, M. (2022). An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection. *Multimedia Systems*, *28*(4), 1309-1323.

Rathore, A., Bhongade, R. A., & Sharma, M. M. (2023). Hybrid neural network for non-image-based knee osteoarthritis prediction. *Multidisciplinary Science Journal*, *5*.

Harangi, B. (2018). Skin lesion classification with ensembles of deep convolutional neural networks. *Journal of biomedical informatics*, *86*, 25-32.

Imran, A., Nasir, A., Bilal, M., Sun, G., Alzahrani, A., & Almuhaimeed, A. (2022). Skin cancer detection using combined decision of deep learners. *IEEE Access*, *10*, 118198-118212.

Alizadeh, S. M., & Mahloojifar, A. (2021). Automatic skin cancer detection in dermoscopy images by combining convolutional neural networks and texture features. *International Journal of Imaging Systems and Technology*, *31*(2), 695-707.

Pacheco, A. G., & Krohling, R. A. (2021). An attention-based mechanism to combine images and metadata in deep learning models applied to skin cancer classification. *IEEE journal of biomedical and health informatics*, *25*(9), 3554-3563.

Codella, N. C., Nguyen, Q. B., Pankanti, S., Gutman, D. A., Helba, B., Halpern, A. C., & Smith, J. R. (2017). Deep learning ensembles for melanoma recognition in dermoscopy images. *IBM Journal of Research and Development*, *61*(4/5), 5-1.

Celebi, M. E., Codella, N., & Halpern, A. (2019). Dermoscopy image analysis: overview and future directions. *IEEE journal of biomedical and health informatics*, *23*(2), 474-478.

Wen, D., Khan, S. M., Xu, A. J., Ibrahim, H., Smith, L., Caballero, J., ... & Matin, R. N. (2022). Characteristics of publicly available skin cancer image datasets: a systematic review. *The Lancet Digital Health*, *4*(1), e64-e74.

Lee, T., Ng, V., Gallagher, R., Coldman, A., & McLean, D. (1997). Dullrazor®: A software approach to hair removal from images. *Computers in biology and medicine*, *27*(6), 533-543.

Michele, A., Colin, V., & Santika, D. D. (2019). Mobilenet convolutional neural networks and support vector machines for palmprint recognition. *Procedia Computer Science*, *157*, 110-117.

Zhu, Y., & Newsam, S. (2017, September). Densenet for dense flow. In *2017 IEEE international conference on image processing (ICIP)* (pp. 790-794). IEEE.

Pang, B., Nijkamp, E., & Wu, Y. N. (2020). Deep learning with tensorflow: A review. *Journal of Educational and Behavioral Statistics*, *45*(2), 227-248.

Ketkar, N., & Ketkar, N. (2017). Stochastic gradient descent. *Deep learning with Python: A hands-on introduction*, 113-132.

Mascarenhas, S., & Agarwal, M. (2021, November). A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. In *2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON)* (Vol. 1, pp. 96-99). IEEE.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).