Accurate Deep Learning Algorithms for Skin Lesion Classification

Sohaib Najat Hasan[ID]

Kirkuk Technical Institute, Northern Technical University, Kirkuk 36001, Iraq

Corresponding Author Email: suhaibnajat@ntu.edu.iq

ABSTRACT

A skin lesion is any irregularity or alteration to the texture, color, or appearance of the skin. It arises from a number of skin illnesses, such as malignancies, autoimmune diseases, allergies, and infections. Early detection and precise diagnosis of skin lesions are crucial for effective treatment and management of these disorders. Dermatologists and other healthcare professionals have traditionally diagnosed skin lesions through visual inspection. However, using this approach might result in a delayed or incorrect diagnosis. Skin lesion categorization accuracy has significantly improved as a result of recent advancements in deep learning techniques. This study looks at the different deep learning techniques used to classify skin lesions. These include transfer learning (DenseNet201 and ResNet52V2) and Convolutional Neural Networks (CNNs). Our study's results show that test images have a 91% accuracy rate, while training images have a 95% accuracy rate.

1. INTRODUCTION

Skin cancer, is a disease that mainly affects men and women equally; however, it is somewhat more common in men. Skin cancer is responsible for 4% of cancer cases in women and 6% of cancer cases in men [1]. It is a major public health concern owing to its rising prevalence and potential severity. Factors leading to the growth in skin cancer cases are [2]:

UV Exposure: Prolonged exposure to ultraviolet (UV) light from the sun or tanning beds increases the chance of developing all forms of skin cancer.

Genetic Predisposition: People who have specific genetic conditions or a family history of skin cancer are more susceptible.

Skin Type: Fair skinned, light-haired, and light-eyed individuals are more vulnerable to UV exposure and ensuing skin cancer.

Immune Suppression: Those who have compromised immune systems—from diseases or medical procedures—are more vulnerable.

Skin lesions refer to a wide range of disorders, ranging from harmless to cancerous, each having unique features and consequences for health. Benign lesions include melanocytic nevi, or moles, as they are more frequently called, which are mostly innocuous but may sometimes undergo a transformation into melanoma. Importantly, melanoma is a type of skin cancer that can be fatal and arises from melanocytes, which are cells that produce color [3]. Past sunburns and excessive sun exposure frequently contribute to its development. For effective melanoma treatment and better patient outcomes, early identification and diagnosis are essential. The information highlights how critical it is to raise awareness, practice prevention, and conduct routine skin exams in order to detect melanoma in its early stages, when it is most curable. Other benign lesions, like dermatofibromas and seborrheic keratoses, typically pose no significant risk, but their removal may occur if they begin to cause symptoms or for aesthetic reasons. Actinic keratoses are precancerous lesions that, if left untreated, may develop into squamous cell carcinoma (SCC). This emphasizes the need for early discovery and treatment. Basal cell carcinoma (BCC), is the most prevalent form of skin cancer in White communities, accounting for 75% of all skin malignancies. Typically, they appear in the head and neck region (w83%) [4]. Squamous cell carcinoma (SCC) The incidence of this type is lower than that of BCC. It grows in arid and rugged regions and between squamous cells. Nevertheless, it may manifest in specific regions of the epidermis that receive an increased amount of radiation. It may manifest as red patches. Similar to BCC, it has the potential to disseminate to other parts of the body; however, certain treatments have been identified to avert its progression [5]. Vascular lesions, such as angiomas and pyogenic granulomas, are noncancerous growths of blood vessels that often appear as red or purple patches on the skin [6].

The artificial neural network is utilized by deep learning to examine data in various categories and uncover patterns to learn from it. This type of learning is based on artificial neural network architecture, which they are built to emulate the structural design of the human brain. These convolutions, are in fact, layers that serve to identify and filter only pertinent information from the data [7].

Deep learning mainly uses artificial neural networks to analyze images of skin disease and detect skin cancer. It has great predictive ability in the diagnosis of skin cancer with a primary role in most malignant cancers [8]. Deep learning, in turn, is a basis of training of artificial neural networks with large data sets.

The challenges facing skin cancer diagnosis techniques are numerous, including the few annotated datasets and the need of significant processing power. Nevertheless, the effectiveness of deep learning relies on process development that has the potential to resolve these problems and improve the diagnostic precision of skin cancer.

Skin lesions are particularly interesting and useful in the context of CNNs to identify and classify the lesions [9]. Convolutional Neural Networks (CNNs) are a type of neural network that are specifically designed for image recognition tasks, and of course take advantage of hierarchial structure that are supposed to abstract important features from the images. CNN is a composition of many layers collaborate with each other in features extraction and classification. The beginning layers of the architecture of a neural network identify basic features of an image such as edges & corners, while the latest layers will pick-up more complex features. E.g. shapes, patterns. The last layer of the network outputs the classification result, that is, the type of skin lesion. This allows the network to learn increasingly complex representations of the input data as it goes through the layers of the network.

CNNs are quite good at correctly categorizing skin lesions—more especially, they are very good at melanoma detection. Unfortunately, there are frequently few annotated datasets available in the dermatological sector, which poses a serious problem because CNNs need a sizable amount of labeled data to train [10].

The application of transfer learning in deep learning is a tactic meant to increase model performance on fewer datasets. By utilizing the knowledge gathered from previously trained models on larger datasets, this is accomplished. This way of classifying skin lesions uses CNN models that have already been trained on large datasets (ImageNet) to pull out important features from images of skin lesions. A smaller CNN model is then trained using these extracted features on a more compact series of dataset of skin lesions [11].

It has been demonstrated that transfer learning increases skin lesion categorization accuracy, especially when there is little labeled data available. However, the pre-trained model selection and the particular architecture of the smaller CNN can have a significant impact on the transfer learning performance.

In spite of the fact that diagnostic technologies have seen some advancements, there are still several obstacles and constraints that remain in the area of skin lesion classification:

Subjectivity and Variability in Visual Inspections: Conventional approaches to identifying skin lesions mainly depend on visual inspections conducted by dermatologists. The subjectivity and variability about the matter may yield inconsistent diagnoses, misdiagnoses and ultimately late-stage delayed treatment.

Data Imbalance: The lack of an even distribution between benign lesion and malignant lesions, makes the varied appearance of skin lesions a real challenge. Such potential outcome could be adversely biased the functioning of a model.

Inadequate Annotated Data: Datasets with high-quality annotated data are very important for the well-fitting of deep learning models. Provided these datasets house the necessary ground truth for supervised learning algorithms to correctly understand and classify various skin lesions. But the scarcity of available information makes it particularly difficult for the medical field to develop and validate reliable diagnosis models.

In this paper, We investigate and conduct a thorough analysis of different deep learning algorithms for skin lesion detection, such as Convolutional Neural Networks and transfer learning models, like DenseNet201 and ResNet52V2. Our results demonstrate that the proposed models achieving 95% accuracy on training images and 91% accuracy on test image, pointing their ability to appear in classifying skin lesion types. This study aims to aid ongoing efforts to improve skin cancer detection and outcomes of patients with Artificial Intelligence technology.

2. LITERATURE REVIEW

Skin cancer classification through image analysis has been developed tremendously over time. In the quest to increase efficacy, the diagnostic method has been explored and evaluated in many ways.

Gessert et al. [12]: "Skin Lesion Diagnosis using Ensembles, Unscaled Multi-Crop Evaluation and Loss Weighting". They employed two built-in datasets for the purpose of their research, which are ISIC 2018 and HAM10000. Different models were experimented by the researchers, such as ResNet50, Densenet121, Densenet161, or Densenet169. Densenet121 achieved an accuracy of 88%, ResNet50 and Densenet161 both achieved 86%, and Densenet169 achieved 85%.

The article "Skin Lesion Classification Using Pre-Trained DenseNet201 Deep Neural Network" authored by Jasil and Ulagamuthalvi [13] present in the 3rd International Conference on Signal Processing and Communication, 2021. They reported good performance at skin lesion classification with DenseNet201 architecture. This work is showing how easily a pre-trained model can improve classification accuracy where they exhibit poor performance for detection of most classes of skin lesions. This model gives 77% of test accuracy and 95% of training accuracy.

Al-Masni et al. [14] developed a classification model to detect skin lesions. This model employs two different deep learning architectures: ResNet-50 and DenseNet-201. They evaluated the model on ISIC 2016, ISIC 2017 and ISIC 2018 datasets. For ISIC 2016, ResNet-50 yielded 79.95% accuracy and DenseNet-201 yielded 81.27% accuracy. ISIC 2017 ResNet-50: 81.57% DenseNet-201: 73.44% Finally, both ResNet-50 and DenseNet-201 yielded higher rate of accuracy (89.28%, 88.70%) respectivly by using ISIC 2018 dataset.

In 2020, Rahman and Ami [15] worked on classification of skin lesions using transfer learning techniques. He experimented with the above approach with the HAM10000 dataset and the ResNet and DenseNet models. Results showed the effectiveness rate of ResNet 87% and dense net showed the high rate of 89%. Following that in 2021 a project lead by Rahman et al. [16], emphasized on multiclass Skin lesion classification. Ensemble learning has been incorporated which utilized the HAM and ISIC 2019 datasets. As previous paper, ResNest and DesensNet models were used. When testing on this dataset, ResNet could achieve an accuracy of 75%, while another deeper network, DenseNet performed much better reaching an accuracy of 84%.

Kondaveeti and Edupuganti [17] used HAM10000 dataset, which consists of 10015 images that are unbalanced from seven different skin lesion categories. Transfer Learning was implemented by researchers on the pre-trained models like ResNet50, InceptionV3, MobileNet, and Xception and then added a specific layers to the classification model task. The class imbalance was handled using data augmentation and

class-weighted loss. ResNet50 was one of the most successful models on the dataset achieving 90% accuracy, 0.89 weighted average precision and a 0.90 recall. It is proven by the research, that using pre-trained model highly predictable the classification accuracy despite very few of data which can be conventional for clinical decision support in dermatology. The researchers believe that this study provides evidence that the combination of deep learning and medical imaging could improve diagnostic process to prevent unneeded biopsies and improve patient outcomes.

Jasil and Ulagamuthalvi [18] employed transfer learning to examine deep learning techniques to classify skin lesions. They performed all experiments using three pre-trained models (Inception V3, VGG16, and VGG19) on the ISIC 2018 dataset that contains 2487 training images and 604 test images in seven classes of skin lesion. Their per-processing method included resizing and centering the image and then data augmentation for the creation of more training images. Pre-trained models were fine-tuned by replacing the final layers with softmax layers for skin lesion classification. Results found that test accuracy of VGG16 was the best (77%) and VGG19 (76%) was next best while Inception V3 (74%) was the lowest in test accuracy. Evaluation of the models were done by accuracy, precision, recall, and F1-score, and classification performance of each class was measured with confusion matrices.

3. METHODOLOGY

3.1 HAM10000 dataset

The HAM10000 [19] is a large dataset of skin lesions acquired for a scientific research purpose from the Department of Dermatology at the Medical University, Vienna. It consists of 10,015 dermoscopic images categorized into seven different groups showing pigmented skin lesions:

(1) (AKIEC) Actinic keratosis / Bowen's disease. the lesion is a non-malignant tumor, but that may evolve into a malignant disease (squamous cell carcinoma).

(2) Malignant - (BCC) Basal cell carcinoma.

(3) Non-malignant – (BKL) keratosis Lesion.

(4) Benign – (DF) Dermatofibroma.

(5) Malignant – (MEL) Melanoma.

(6) Benign (non-cancerous) - (NV) Nevi Melanocytic.

(7) (VASC) Vascular lesions: can be either Malignant or Non-malignant lesion.

Figure 1 shows several representative images of skin lesions. The collection consists of images with varying resolutions. Because to resolution variability, categorization of skin lesions becomes challanging task.

3.2 Pre-process methodology

To improve the quality of images and ensure their compatibility with the requirements of neural networks used in classification and pattern recognition. The dataset is organized into six folders and known as class names. Because the raw dataset is highly imbalanced, data balancing was performed using down sampling method. The dataset is then divided into two sets: 20% for testing and 80% for training. The training and test sets were then subjected to image preparation methods, which included image normalization the augmentation. Figure 2. illustrate the main phases of preprocessing method.
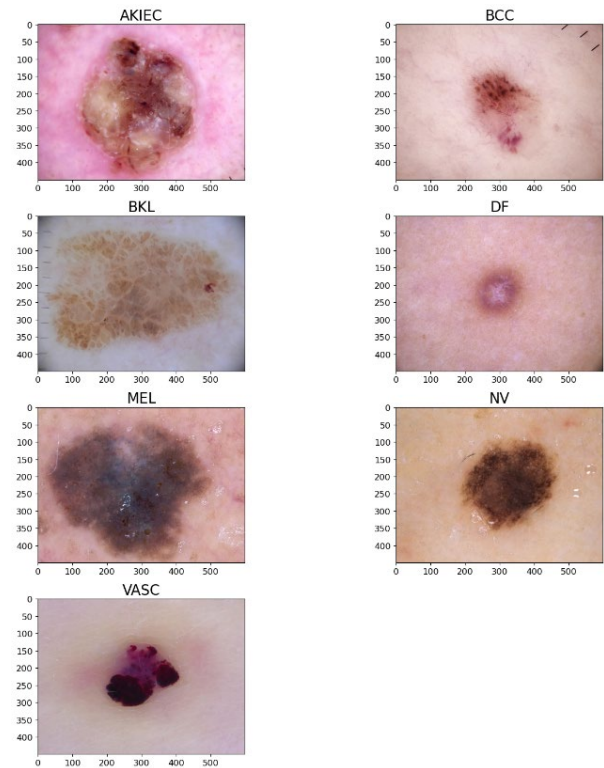

Figure 1. Sample photos from the dataset

Preprocessing of HAM10000 dataset images was applied in the following steps

First, resizing images to fixed dimensions is a crucial step to feed the neural network. All images in the dataset were resized to 224×224 pixels, which is the optimal size for many pre-trained neural networks such as ResNet and DenseNet. This change ensures that all images enter the network with the same dimensions, reducing the complexity of the learning process and increasing training efficiency [20].

Second, image balancing: The dataset is highly imbalanced, with the (Melanocytic nevi) class representing the majority of the images (31.2%), and the other classes having much fewer samples (akiec: 6.8%, bcc: 10.7%, bkl: 22.8%, df: 2.4%, mel: 23.1%, vasc: 3.0%) shown in Figure 3. The images are of various sizes and resolutions, and they were acquired from a range of sources and conditions, resulting in variability in terms of lighting, focus, and quality.

To address class imbalance in datasets, one way is to resample the training dataset randomly. Undersampling entails removing instances from the majority class, and oversampling [21] entails copying examples from the minority class. Three phases made up the oversampling technique utilized in this study:

1) First, a subset of the HAM10000 dataset that contains only the "Melanocytic nevi" class is created.

2) The subset is then randomly sampled to reduce the number of samples in the "Melanocytic nevi" class to 1500.

3) Next, the remaining classes in the HAM10000 dataset are oversampled to balance the number of samples across all classes. This can be achieved using techniques such as random oversampling, which involves randomly duplicating samples from the minority classes until the number of samples in all classes is equal. The result is shown in Figure 4.
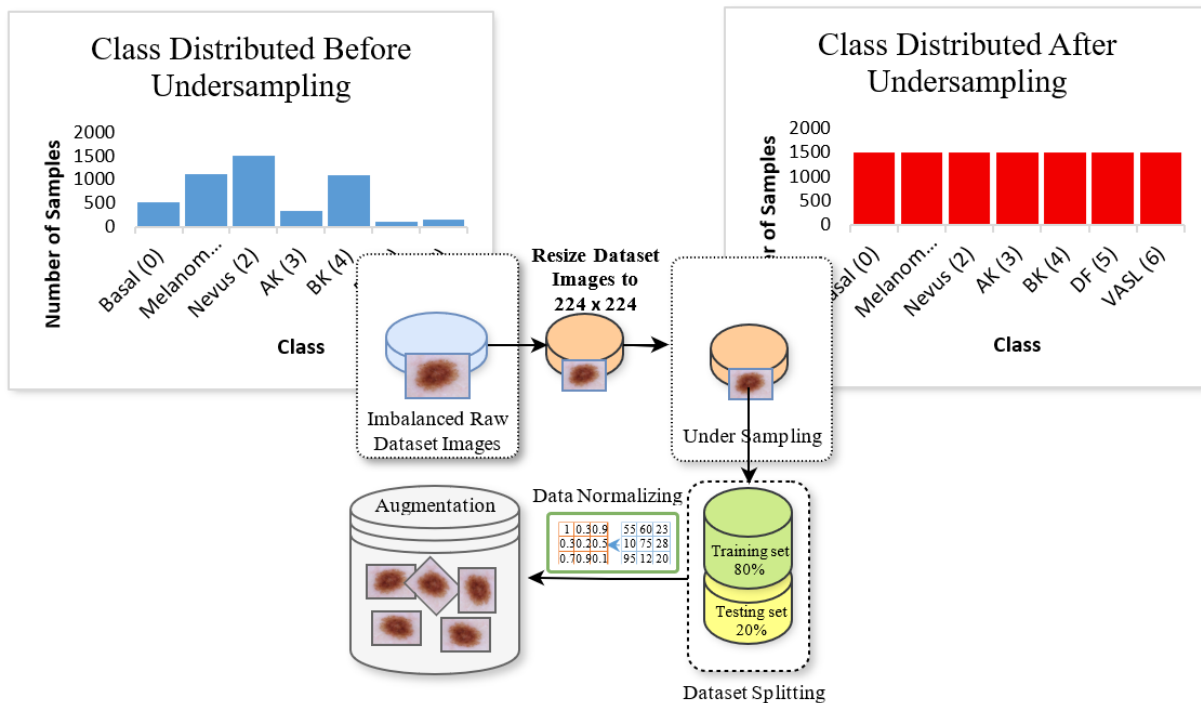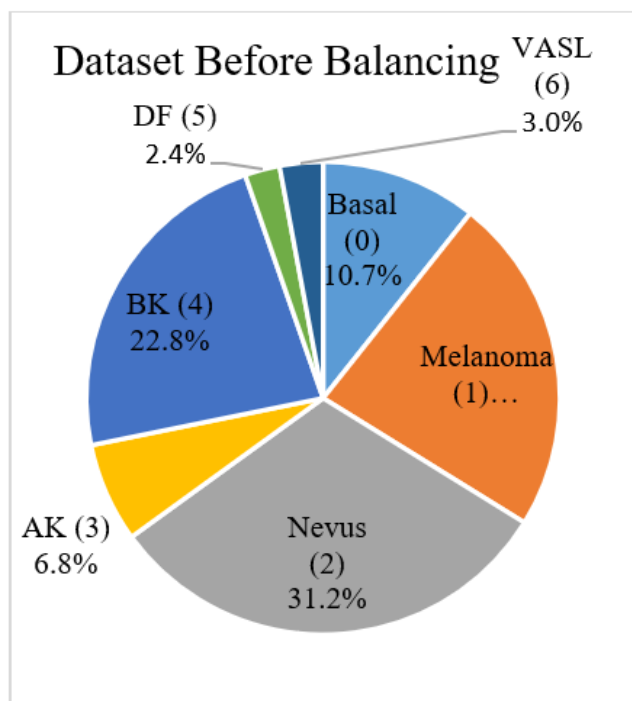
Figure 2. Preprocessing phases

Figure 3. Database before balancing

out, flipping horizontally and vertically [24].

By performing these four steps: Image Resizing, Image Balancing, Pixel Scale Adjustment, and, Image Augmentation which speeds up the neural network's training process and contributes to improve the model's performance and increasing the classification accuracy.
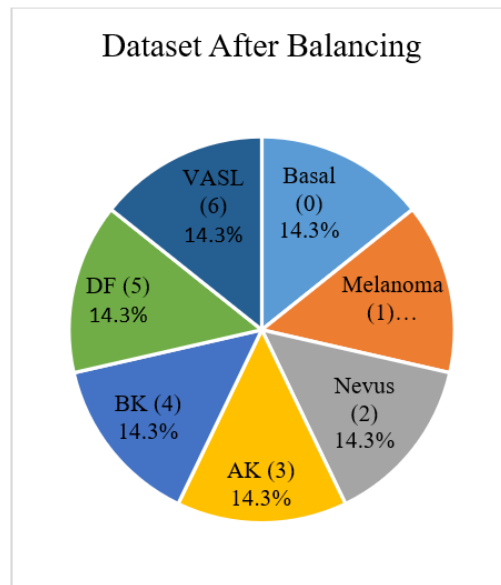

Figure 4. Database after balancing

3.3 DenseNet201

DenseNet201 is a Convolutional Neural Network architecture that was introduced in 2018 as an extension to the original DenseNet architecture. DenseNet201 gets its name from the sum of 201 layers that it contains [25].

DenseNet201, like the original DenseNet architecture, makes use of a densely connected network structure in which all layers above it provide input to the layer below it,

Third, normalizing (adjusting the pixel values to range from 0 to 1): adjusting is done by dividing the value of each pixel by 255 (the maximum pixel value in grayscale images or color images). This uniform range makes it easier for the neural network to process images better and ensures the stability of the input values [22, 23].

Fourth: image augmentation: When the number of images has decreased due to the use of image balancing, a technique is used to increase the diversity of images available to train deep learning models without actually collecting new images. This was achieved by applying different transformations to the HAM10000 dataset images, such as rotating, zooming in and

encouraging feature reuse and lowering the number of parameters.

By adding bottleneck layers this reduces the computing cost of the network. They use 1x1 convolutions to reduce the number of input channels before the main convolution operation is performed. They then use another1x1 convolution to go back to the original output channels. Meaning, the network works faster and more efficient and needs to learn fewer parameters during the convolution process.

Moreover, DenseNet201 utilizes feature concatenation which concatenates feature maps from numerous layers before they proceed to the next layer. This is because the network has higher precision and a better feature acceptance mode for low-level and high-level features.

Overall, still DenseNet121 is the best deep learning framework for image classification tasks. It has high use in computer vision mainly used is in medical image analysis.

3.4 ResNet152V2

Is a type of deep Convolutional Neural Network structure. It has shown great success in the field of image classification, such as the classification of skin lesions. This is an extension of the ResNet family of architecture which used residual blocks to train deeper networks to prevent vanishing gradient problem [26].

ResNet152V2 is composed of 152 layers, most of which are convolutional [27]. The model contains four stages in its architecture, with the count of residual blocks being different throughout them. First stage 7 residual layers, second stage 35 residual layers, third stage 8 residual layers, final stage 3 residual blocks.

Each block is composed of two or three convolutional layers followed by ReLU activations and batch normalization. The network can learn residual functions that can be used to modify the input features by adding the output of the convolutional layers to the input of the residual block. The residual blocks also include skip connections, which enable gradient flow through the network during backpropagation and prevent the vanishing gradient problem.

ResNet152V2 has global average pooling in addition to convolutional layers and residual blocks. This decreases the spatial dimensions of the feature maps to one value per feature map. After that, a fully connected layer processes the generated feature vector to provide the output that is needed to complete the classification task.

ResNet152V2 has been shown to be highly effective in skin lesion classification, achieving high accuracy on various datasets including the HAM10000 dataset.

3.5 Metrics

Actual Values

Figure 5. Confusion matrix [28, 29]

Figure 5 illustrates confusion matrix which is utilized to valuate binary classification models by representing True (1) or False (0) for the actual classes and Positive (1) or Negative (0) for the resulting or predicted classes.

True Positive (TP): The predictive model precisely estimates the positive class.

False Positive (FP): The predictive model inaccurately predicts the positive class (also called a Type II error).

True Negative (TN): The predictive model accurately predicts the negative class.

False Negative (FN): The predictive model inaccurately estimated the negative class (also called a Type II error) [30].

Several important measurements may be derived from the confusion matrix to assess the classification model's performance:

Accuracy:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

Accuracy is the percentage of accurately predicted cases (both positive and negative) out of all instances.

Precision (Positive Predictive Value):

$$Precision = \frac{TP}{TP+FP} \qquad (2)$$

Precision refers to the fraction of genuine positive predictions among all positive predictions produced by the model.

Recall (Sensitivity or True Positive Rate):

$$Recall = \frac{TP}{TP+FN} \qquad (3)$$

Recall calculates the fraction of genuine positive cases among all actual positive instances.

Specificity (True Negative Rate):

$$Specificity = \frac{TN}{TN+FP} \qquad (4)$$

Specificity refers to the fraction of real negative cases among all actual negative instances.

F1 Score:

$$F1\ Score = 2 \times \frac{Presision \times Recall}{resision+Recall} \qquad (5)$$

The F1 score is the combined mean of precision and recall, resulting in a balance of the two measurements.

Receiver Operating Characteristic (ROC) Curve:
The ROC curve shows how a binary classifier's diagnostic performance changes when the discrimination threshold is adjusted. It compares the actual positive rate (sensitivity or recall) with the false positive rate (1-specificity).

3.6 Proposed model

The proposed model to classify skin lesions involved several key steps as it shown in Figure 6. In order to match the input picture dimensions for the selected models, Densenet201 and ResNet152V2, a preprocessing step was utilized to scale the images to a consistent size of 224 × 224 pixels. Oversampling was used to address the issue of class imbalance

in the HAM10000 dataset. To equalize the distribution of samples across all classes, more members of the minority class have to be created.

Using the preprocessed and oversampled HAM10000 dataset, the Densenet201 model was used in the study's initial phase, and it produced an accuracy of 88%. Using a pre-trained Densenet121 model—which attained an accuracy of 86%—was the second phase. Using a pre-trained Resnet152V2 model, the third stage resulted in an accuracy rate of 88%.

To further improve accuracy, a combined analysis of the output from the top-performing models was done.

In order to optimize the performance of the suggested model, the outputs of the Densenet201 and Resnet152V2 models were integrated by combining them, and then the following layers were added:

1. The Global Average Pooling Layer processes the averages across the whole spatial dimension of the feature maps after receiving the combined model outputs. As a result, a global feature vector is produced, which highlights the salient characteristics of the input image.

2. The addition of the dense layer, which has 256 neurons and a ReLU activation function, enables the network to learn more intricate features from the input data.

3. Dropout 0.2: During training, this layer arbitrarily removes 20% of the activations from the preceding dense layer. This regularization method aids in avoiding overfitting.

4. Dense layer (64, activation='relu'): This layer adds another dense layer with 64 neurons and applies the ReLU activation function. This further allows the network to learn more complex features of the input data.

5. The second Dropout layer randomly drops out 20% of the activations from the previous Dense layer during training.

6. The final output layer of the model, the Dense layer with 7 neurons, is included, one for each class in the dataset. Over the classes, a probability distribution is produced using the softmax activation function. For the input image, the predicted class is determined by selecting the class with the highest probability.

We used the categorical cross-entropy loss and the Adam optimizer for training the model. The model was trained for 60 epochs with the batch size of 64. The model was checkpointed with the best model weights based on validation loss. In order to prevent overfitting and increase the variability of the training data, image augmentation like (rotation, zoom of the images, changing the width and heigh shift range, flipping) were used randomly to train the images during the training. Lastly, a series of metrics, including the F1-score, Precision, Recall, AUC-ROC, and Classification Accuracy, were used to evaluate the performance of the recommended model on the test set.

4. RESULTS AND DISSCUTION

In general, results come from a validation dataset, which are 20% of all images in the HAM10000 dataset, well shuffled. The dataset used consists of a large sample of all images from seven different types of cancer skin. The models, DenseNet201 and ResNet152V2 were trained on a local GPU (NVIDIA GEFORCE GTX 1070), and were implemented through the Keras library to achieve the observed accuracies.

They compared the models on a number of criteria, with accuracy being reported as the primary metric. The accuracy is the percentage of the testing dataset that has been rightly

classified out of all occurrences. In this case, DenseNet201 detected 88% of the skin lesions correctly and had an accuracy of 88% on the testing dataset. Similarly, ResNet152V2 also gained a similar level of performance to DenseNet201, with accuracy of 88.

Serval criteria was uesed to evaluate how the models performed, accuracy is being as the primary metric. The accuracy is the percentage of the testing dataset that has been rightly classified out of all occurrences. DenseNet201 detected 88% of the skin lesions correctly and had an accuracy of 88% on the testing dataset. Similarly, ResNet152V2 also gained a similar level of performance to DenseNet201, with accuracy of 88.

DenseNet201 and ResNet152V2 were combined to form DenseNet201-ResNet152V2, which reported a significant improvement in accuracy. This mixture of the two models means that we can leverage the unique attributes and properties that each architecture has learned. This highest accuracy of 91% is better than the individual performance of DenseNet201 and ResNet152V2. It is useful for getting the counts of true positive, true negative, false positive, and false negative predictions for each class given a confusion matrix which also show how the model is classifying images of different classes.

4.1 Confusion matrix of models' comparison

Figures 7-9 show detailed statistics in the confusion matrix relating to how well the DenseNet201, ResNet152V2, and the combined DenseNet201-ResNet152V2 models performed at classifying skin lesion images of the different types of skin lesion; it can be seen that the three models can well distinguish among the skin lesions. This section compares the models which shed light on the respective strengths and weakness of each model.

DF and VASC display perfect true positive rates for all three models, showing the robustness of lesions detection. The combined model shows improvement in true positive rates for BCC (281), Melanoma (225), Nevus (256), AK (285), and BKL (255) compared to the individual models.

DenseNet201 has fewer misclassifications overall compared to ResNet152V2, especially for Melanoma and Nevus. ResNet152V2 shows higher misclassification rates for Nevus and Melanoma, indicating difficulty in distinguishing these classes. In most cases (BCC, Melanoma, Nevus, and AK), the misclassifications are reduced by combining DenseNet201 and ResNet152V2, showing the advantages on combining features from the both architectures.

BCC (Basal Cell Carcinoma): The composite model has the utmost true positives and the least number of misdetections out of the three.

Melanoma: True positive rate is better and number of false positives is less for the combined model as compared to DenseNet201 as well as ResNet152V2, thus the combined model achieves better detection accuracy.

Nevus: The combined model offers an optimal trade-off between the true positive rate and the misclassification, and provides a better performance than the individual models.

AK (Actinic Keratosis): This combined model displays the highest true positive rate along with fewest number of misclassifications, which denotes the best performance of the model.

The multiple skin lesion types confusion matrices show improved classification when combining DenseNet201 and

ResNet152V2. Our combined model performs with better sensitivity and less misclassifications overall, especially for difficult classes such as Melanoma and Nevus. The improvements provide an illustrative example of the benefit of using strengths of different models to improve the performance in accurate and reliable skin lesion classification in clinical use.
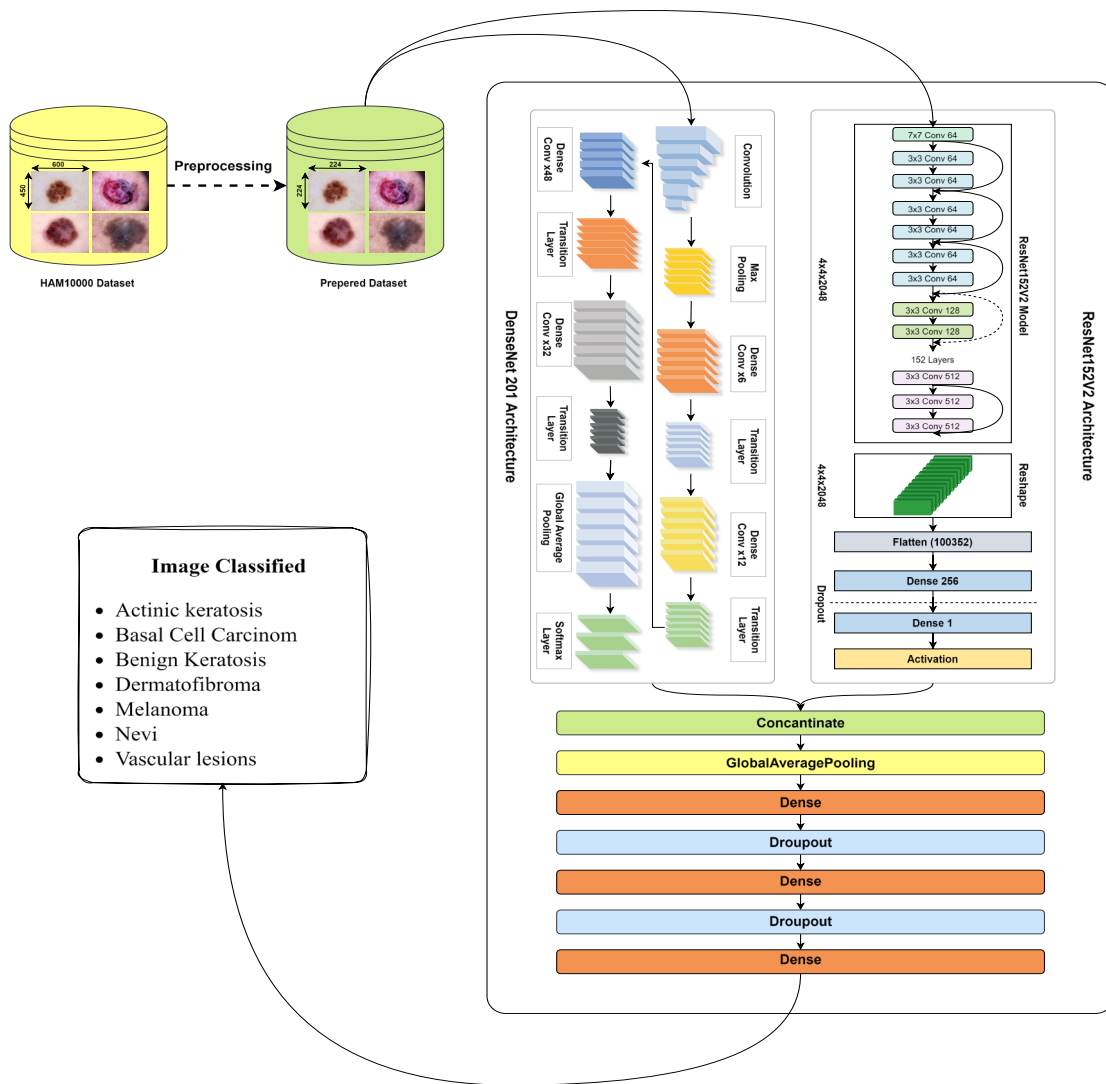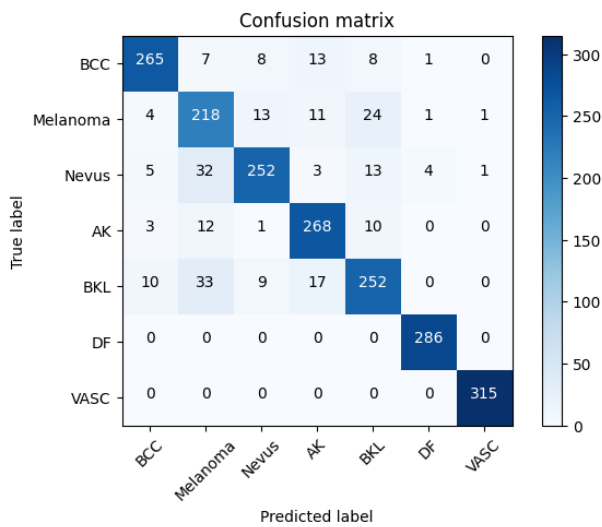

Figure 6. Structure of the suggested model
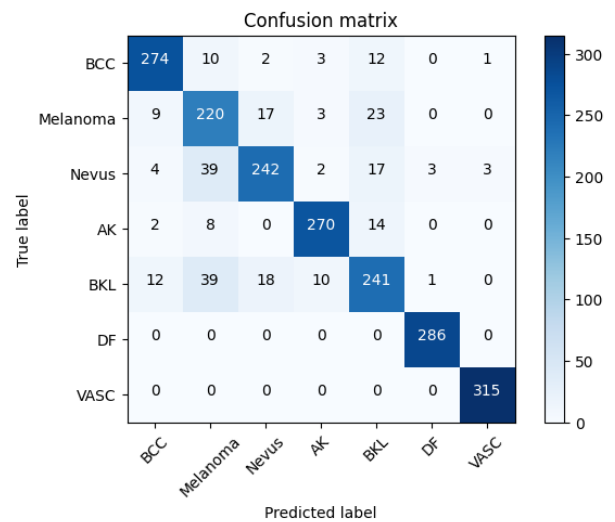

Figure 7. DenseNet201 confusion matrix

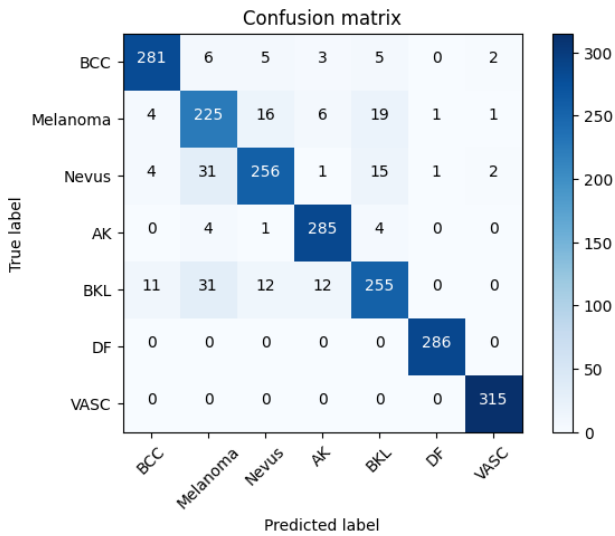**Figure 8.** ResNet152V2 confusion matrix

Figure 9. DenseNet201- ResNet152V2 confusion matrix

4.2 Evaluation of training and validation performance

Figure 10 demonstrate the testing and training accuracy curve over 60 epochs. During the Initial Training Phase (0 - 10 epochs), training accuracy and testing accuracy starts increasing rapidly which illustrates learning happens efficiently. There will be slight fluctuations in testing accuracy as the model trains. In the middle training phase (10-20 epochs), training accuracy improves and stabilizes around the 20th epoch. Validation accuracy also stabilizes, slightly lower than training accuracy, indicating good generalization without significant overfitting. In the later training phase (20-60 epochs), Training accuracy plateaus above 0.9, showing the model has learned the patterns well. Validation accuracy remains around 0.9, closely aligning with training accuracy, suggesting strong generalization and minimal overfitting.
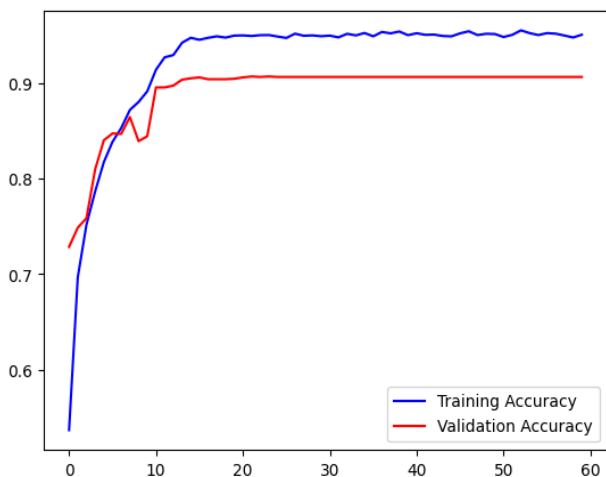

Figure 10. Validation ACC and Training ACC performance of (DenseNet201-ResNet152V2) model

Figure 11 illustrates the training and validation loss curves over 60 epochs. In the initial training phase (0-10 epochs), Both training and validation loss decrease rapidly, reflecting effective learning and error reduction. Slight fluctuations in validation loss are expected as the model fine-tunes its parameters. In the middle training phase (10-20 epochs), Training loss continues to decline and stabilizes around the

20th epoch. Validation loss also decreases but remains slightly higher than training loss, indicating a small generalization error. In the later training phase (20-60 epochs), training loss plateaus at a low value, showing minimized error on the training set. Validation loss remains stable and consistently higher than training loss, indicating good generalization and minimal overfitting.

The training and validation metrics demonstrate that the model is well-trained, with effective learning and minimal overfitting. The early convergence and stability of accuracy and loss curves suggest that the model efficiently captures the data's patterns, making it robust for practical applications in skin lesion classification. The close alignment of training and validation accuracy, along with stable validation loss, indicates good generalization to unseen data with minimal overfitting.
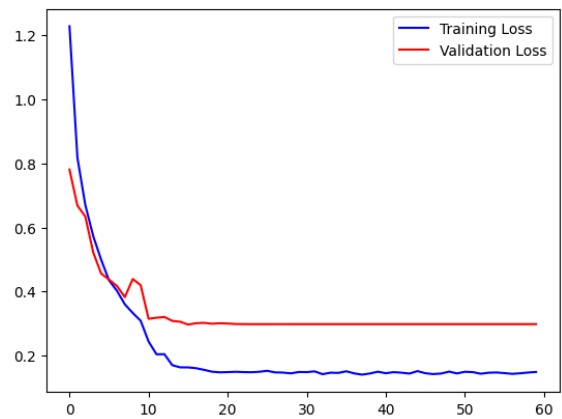

Figure 11. Validation loss and Training loss performance of (DenseNet201-ResNet152V2) model

4.3 Evaluation of model performance

This section evaluates several deep learning models and how they perform on different datasets. It emphasizes the impact of model architecture and transfer learning on classification accuracy. Table 1 provides a comparison between our proposed approach and existing models.

The accuracy of ResNet models exhibits significant variation, ranging from 75% to 90%. This suggests that the dataset and implementation specifics heavily influence their performance. In research that used transfer learning, ResNet50 demonstrated its resilience by achieving a peak accuracy of 90%. This indicates that the model's pre-trained weights were effectively fine-tuned for individual datasets. The different levels of accuracy (79.95%, 81.57%, and 89.28%) seen in ResNet-50 tests show how dataset properties and preprocessing methods can affect how well the model works.

DenseNet models provide robust and consistent performance, with DenseNet121 obtaining an accuracy of 88%, DenseNet161 achieving 86%, and DenseNet169 achieving 85%. In a prior study, the DenseNet201 model demonstrated a significant variation in performance, with a training set accuracy of 95% and a test set accuracy of 77%. This disparity indicates the possibility of overfitting, whereby the model exhibits remarkable performance on the training.

Xception demonstrates a competitive accuracy of 84%, positioning it as a feasible option but somewhat lower than the high-performing DenseNet and ResNet models.

InceptionV3 has varying performance, achieving accuracies of 85% and 74%. This means that applying it to a

variety of datasets compromises its consistency or necessitates further fine-tuning.

MobileNet achieves a commendable accuracy of 87%, making it highly suitable for applications that need efficient and lightweight models.

Compared to more modern and deep networks like DenseNet and ResNet, VGG16 and VGG19 show lower accuracies of 77% and 76%, respectively.

Combining models generally allows for greater accuracy. The ensemble of DenseNet201 and ResNet152V2 demonstrated this by achieving 91% accuracy. That implies even better overall performance can be achieved by combining the strengths and overcoming the weaknesses from several architectures. Combining the ResNet and DenseNet models implies that both architectures enhance each other, which may compensate for any shortcomings in the individual models and make the classification task more accurate.

Table 1. Comparing the suggested model with the literature

Reference	Used dataset	Model	Accuracy
[12]	HAM1000 + ISIC 2018	ResNet50	86
		Densenet121	88
		Densenet161	86
		Densenet169	85
[14]	ISIC 2016	ResNet-50	79.95
		DenseNet-201	81.27
	ISIC 2017	ResNet-50	81.57
		DenseNet-201	73.44
	ISIC 2018	ResNet-50	89.28
		DenseNet-201	88.70
[13]	ISIC 2018	DenseNet201	95% of training set 77% of test set
[15]	HAM10000	ResNet	87
		DenseNet	89
[16]	HAM + ISIC 2019	ResNet	75
		DenseNet	84
		Xception	84
[17]	HAM10000	InceptionV3	85
		MobileNet	87
		ResNet50	90
[18]	ISIC 2018	Inception V3	74
		VGG19	76
		VGG16	77
Our work	HAM10000	DenseNet201	88
		ResNet152V2	88
		DenseNet201-ResNet152V2	91

4.4 Innovative points and future research

The proposed skin lesion classification model has novel characteristics including the native image resizing to 224 × 224 as a sole preprocessing step, an oversampling strategy to handle class imbalance, and a sequential use of Densenet201, Densenet121 and Resnet152V2 models for the different capture the advantages of different feature extraction models. To increase the classification accuracy and robustness, we ensemble the DenseNet201 and ResNet152V2 outputs in the model with an ensemble operation. Moreover, using the pre-trained model illustrates the potential of transfer learning in improving diagnostic performance with limited annotated data.

To counteract data being unbalanced and insufficient, it might actually be more interesting for future work to study state-of-the-art techniques using data augmentation such as Generative Adversarial Networks (GANs). In addition, while

imagery on its own is effective for identifying skin lesions, not all available information is captured. Other digital data, such as age, sex, community, and medical history might serve to improve model accuracy. The additional data would provide more context to perhaps reveal patterns or variations not otherwise visually obvious. In addition, hybrid models using multimodal data with deep learning in collaboration with other machine learning algorithms may enhance the diagnosis accuracy and overall in-depth view of categorical classification of skin lesion. The greatest challenge in building competent models is the lack of rich and full data thanks to which to train such models, so future research needs to focus on collecting data from different ages, sexes, group and settings to make sure the models can be certified to perform in all situations. Another critical aspect when developing a machine learning tool is the quality of the images as well as of the supporting data to create reliable and trustworthy training models.

5. CONCLUSIONS

In summary, Convolutional Neural Networks (CNNs) and transfer learning are powerful deep learning techniques that have demonstrated promise of accurate skin lesion classification. Each of these approaches has its own benefits and disadvantages but in speed and accuracy, they have already outpaced traditional methods for skin lesion classification. As deep learning continues to evolve, the ability to accurately and efficiently classify skin lesions should see significant improvement.

However, we need much more research to overcome challenges associated with deep learning in dermatology, for example the lack of labelled datasets and the high computational infrastructure needed. For greater depth and context, we believe that future work should focus on increasing the complexity of CNNs and transfer learning architectures, increasing the volume and quality of labeled datasets, and developing more user-friendly and interpretable solutions for dermatologists and other healthcare workers.

Customized to this learning style, deep learning models can be a game-changer transforming the domain of dermatology, harnessing superior classification accuracy and faster pace of diagnosing skin lesions. Deep learning provides a rich field of study in regard to detect skin cancer and treatments.

Limitations and Challenges:

All deep learning techniques have their own issues and challenges. Dermatology Domain has an Issue with the requirement of large volumes of labeled data for training CNN The interesting architecture from the smaller CNN and the choice of pre-trained model can significantly be the deal breaker on transfer learning. Additionally, both approaches can be slow and require strong computing resources.

REFERENCES

[1] American Cancer Society. (2023). Cancer Facts & Figures 2023. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/2023-cancer-facts-figures.html/, accessed on Jul. 09, 2023.

[2] Khan, N.H., Mir, M., Qian, L., Baloch, M., Khan, M.F.A., Ngowi, E.E., Wu, D.D., Ji, X.Y. (2022). Skin cancer biology and barriers to treatment: Recent applications of

polymeric micro/nanostructures. Journal of Advanced Research, 36: 223-247. https://doi.org/10.1016/j.jare.2021.06.014

[3] Voss, R.K., Woods, T.N., Cromwell, K.D., Nelson, K.C., Cormier, J.N. (2015). Improving outcomes in patients with melanoma: Strategies to ensure an early diagnosis. Patient Related Outcome Measures, 6: 229-242. https://doi.org/10.2147/PROM.S69351

[4] Chummun, S., McLean, N.R. (2017). The management of malignant skin cancers. Surgery (Oxford), 35(9): 519-524. https://doi.org/10.1016/j.mpsur.2017.06.013

[5] Kaldor, J., Shugg, D., Young, B., Dwyer, T., Wang, Y.G. (1993). Non-melanoma skin cancer: Ten years of cancer-registry-based surveillance. International Journal of Cancer, 53(6): 886-891. https://doi.org/10.1002/ijc.2910530603

[6] Hunt, S.J., Santa Cruz, D.J. (2004). Vascular tumors of the skin: A selective review. In Seminars in Diagnostic Pathology, 21(3): 166-218. https://doi.org/10.1053/j.semdp.2005.01.001

[7] Kadampur, M.A., Al Riyaee, S. (2020). Skin cancer detection: Applying a deep learning based model driven architecture in the cloud for classifying dermal cell images. Informatics in Medicine Unlocked, 18: 100282. https://doi.org/10.1016/j.imu.2019.100282

[8] Dildar, M., Akram, S., Irfan, M., Khan, H.U., Ramzan, M., Mahmood, A.R., Alsaiari, S.A., Saeed, A.H.M., Alraddadi, M.O., Mahnashi, M.H. (2021). Skin cancer detection: A review using deep learning techniques. International Journal of Environmental Research and Public Health, 18(10): 5479. https://doi.org/10.3390/ijerph18105479

[9] Convolutional Neural Network in the Medical Imaging. https://encyclopedia.pub/entry/42866/, accessed on Jul. 19, 2023.

[10] IBM. What are Convolutional Neural Networks? https://www.ibm.com/topics/convolutional-neural-networks/, accessed on Jul. 19, 2023.

[11] Weiss, K., Khoshgoftaar, T.M., Wang, D. (2016). A survey of transfer learning. Journal of Big Data, 3: 1-40. https://doi.org/10.1186/s40537-016-0043-6

[12] Gessert, N., Sentker, T., Madesta, F., Schmitz, R., Kniep, H., Baltruschat, I., Werner, R., Schlaefer, A. (2018). Skin lesion diagnosis using ensembles, unscaled multi-crop evaluation and loss weighting. arXiv preprint arXiv:1808.01694. https://doi.org/10.48550/arXiv.1808.01694

[13] Jasil, S.G., Ulagamuthalvi, V. (2021). Skin lesion classification using pre-trained DenseNet201 deep neural network. In 2021 3rd International Conference on Signal Processing and Communication (ICPSC), pp. 393-396. https://doi.org/10.1109/ICSPC51351.2021.9451818

[14] Al-Masni, M.A., Kim, D.H., Kim, T.S. (2020). Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification. Computer Methods and Programs in Biomedicine, 190: 105351. https://doi.org/10.1016/j.cmpb.2020.105351

[15] Rahman, Z., Ami, A.M. (2020). A transfer learning based approach for skin lesion classification from imbalanced data. In 2020 11th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, pp. 65-68. https://doi.org/10.1109/ICECE51571.2020.9393155

[16] Rahman, Z., Hossain, M.S., Islam, M.R., Hasan, M.M., Hridhee, R.A. (2021). An approach for multiclass skin lesion classification based on ensemble learning. Informatics in Medicine Unlocked, 25: 100659. https://doi.org/10.1016/j.imu.2021.100659

[17] Kondaveeti, H.K., Edupuganti, P. (2020). Skin cancer classification using transfer learning. In 2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI), Buldhana, India, pp. 1-4. https://doi.org/10.1109/ICATMRI51801.2020.9398388

[18] Jasil, S.G., Ulagamuthalvi, V. (2021). Deep learning architecture using transfer learning for classification of skin lesions. Journal of Ambient Intelligence and Humanized Computing, 1-8. https://doi.org/10.1007/s12652-021-03062-7

[19] Tschandl, P., Rosendahl, C., Kittler, H. (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific Data, 5(1): 1-9. https://doi.org/10.1038/sdata.2018.161

[20] Younis, H., Bhatti, M.H., Azeem, M. (2019). Classification of skin cancer dermoscopy images using transfer learning. In 2019 15th International Conference on Emerging Technologies (ICET), Peshawar, Pakistan, pp. 1-4. https://doi.org/10.1109/ICET48972.2019.8994508

[21] Bedeir, R.H., Mahmoud, R.O., Zayed, H.H. (2022). Automated multi-class skin cancer classification through concatenated deep learning models. IAES International Journal of Artificial Intelligence, 11(2): 764. https://doi.org/10.11591/ijai.v11.i2.pp764-772

[22] Fadil, O., Abdulah, D. (2023). Skin cancer classification using CNN skin cancer classification using CNN view project.

[23] Islam, M.K., Ali, M.S., Ali, M.M., Haque, M.F., Das, A.A., Hossain, M.M., Duranta, D.S., Rahman, M.A. (2021). Melanoma skin lesions classification using deep Convolutional Neural Network with transfer learning. In 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA), Riyadh, Saudi Arabia, pp. 48-53. https://doi.org/10.1109/CAIDA51941.2021.9425117

[24] Lee, K.W., Chin, R.K.Y. (2020). The effectiveness of data augmentation for melanoma skin cancer prediction using Convolutional Neural Networks. In 2020 IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia, pp. 1-6. https://doi.org/10.1109/IICAIET49801.2020.9257859

[25] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4700-4708. https://doi.org/10.48550/arXiv.1608.06993

[26] He, K., Zhang, X., Ren, S., Sun, J. (2016). Identity mappings in deep residual networks. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, 9908: 630-645. https://doi.org/10.1007/978-3-319-46493-0_38

[27] Kousalya, K., Krishnakumar, B., Aswath, A.S., Gowtham, P.S., Vishal, S.R. (2021). Terrain identification and land price estimation using deep

learning. In AIP Conference Proceedings, 2387(1): 140030. https://doi.org/10.1063/5.0068625

[28] Vujović, Ž. (2021). Classification model evaluation metrics. International Journal of Advanced Computer Science and Applications, 12(6): 599-606. https://doi.org/10.14569/IJACSA.2021.0120670

[29] Wu, M.T. (2022). Confusion matrix and minimum cross-entropy metrics based motion recognition system in the classroom. Scientific Reports, 12(1): 3095. https://doi.org/10.1038/s41598-022-07137-z

[30] Amani, M.A., Marinello, F. (2022). A deep learning-based model to reduce costs and increase productivity in the case of small datasets: a case study in cotton cultivation. Agriculture, 12(2): 267. https://doi.org/10.3390/agriculture12020267