International Journal of Computer Applications (0975 – 8887)
Volume 186 – No.18, April 2024

Skin Lesion Prediction from Dermoscopic Images using Deep Learning

Nazma Hossen Nishat
Dept. of Computer Science and Engineering
Port City International University
Chittagong, Bangladesh

Pranta Paul
Dept. of Computer Science and Engineering
Port City International University
Chittagong, Bangladesh

Farzina Akther
Dept. of Computer Science and Engineering
Port City International University
Chittagong, Bangladesh

Tahmina Akter
Dept. of Computer Science and Engineering
Port City International University
Chittagong, Bangladesh

Muhammad Anwarul Azim
Dept. of Computer Science and Engineering University of Chittagong
Chittagong-4331, Bangladesh

ABSTRACT
Skin lesions, which comprise a wide range of irregularities in skin appearance, might serve as precursors of skin cancer due to the complex interaction of hereditary variables and longterm UV ex- posure. Significant advances in dermatology have been made with the use of deep learning models, notably convolutional neural net- works (CNNs). These models excel in analyzing dermatoscopic pictures, allowing for early and accurate identification of a vari- ety of skin problems. In this work, a complete evaluation of deep learning models for predicting skin lesions is conducted, with an emphasis on accuracy. Notable performers include DenseNet169 and ResNet101, both of which achieve an outstanding 91% accu- racy. Furthermore, a hybrid model obtains an accuracy of 89%, in- dicating its capacity to recognize complicated visual patterns. The study investigates model fusion strategies to capitalize on possible synergy in prediction skills, ultimately improving automated der- matological diagnosis systems. Notable models are DenseNet121, ResNet-50V2, and InceptionResNetV2, which contribute consider- ably with accuracies of 91%, 89%, and 85%, respectively, while MobileNetV2 and VGG-16 provide accuracies of 82% and 80%. These advances, taken together, enable the development of strong and accurate diagnostic technologies capable of efficiently expedit- ing skin health interventions.

Keywords
CNN,Transfer Learning, Data Balancing, Augmentation, Hybird Model

1. INTRODUCTION
The area of dermatology has seen dramatic advances in skin le- sion prediction, owing to the use of deep learning (DL) approaches, notably those based on dermoscopic pictures. Skin lesions, which vary in appearance, frequently cause concerns owing to their po- tential link with skin cancer. Recognizing the significance of early detection, academics have embraced deep learning (DL), specifi- cally convolutional neural networks (CNNs), to improve diagnos- tic capabilities. These models excel in analyzing dermoscopic pic- tures, allowing for accurate detection and categorization of numer- ous skin conditions. The predictive value of these models lies from their capacity to detect minor patterns and features associated with certain lesions, allowing for early intervention. Automated predic- tion of skin lesions utilizing deep learning models, such as CNN, represents a paradigm leap in dermatological diagnostics by dra- matically lowering analysis time while also giving a reliable and accurate way of identifying possible issues. This invention not only complements existing diagnostic approaches, but also has the po- tential to improve patient outcomes through prompt treatments. As deep learning advances, a thorough evaluation of various mod- els, such as DenseNet169, ResNet101, and hybrid combinations, on distinct dermoscopic datasets becomes increasingly important. Such assessments, which take into account performance measures as accuracy, not only highlight the strengths of certain models but also direct the investigation of model fusion approaches for possi- ble synergies. In summary, the combination of deep learning with dermoscopic imaging ushers in a new era in dermatology, promis- ing improved predictive capacities and considerably contributing to the early identification and management of skin diseases.

2. LITERATURE REVIEW
According to the research of Neeshma, A. et al., [1] multi-class skin lesions may be classified into seven groups using a deep learning technique that uses the DenseNet-121 architecture and transfer learning. With an accuracy of 82.1%, the recommended model finished the classification task. In order to avoid bias towards a class with a greater picture count, the authors emphasise how important it is to employ balanced datasets in classification tasks. The study shows how promising current technology is for treating and diagnosing skin cancer. Taken together, the proposed approach provides a viable approach to accurate skin lesion classification. In ”Deep Learning-Based Classification of Dermoscopic Images for Skin Lesions” by Ahmet Furkan Sonmez et al., [2] a deep learning-based approach using the MobileNetV2 model is pro- posed to classify skin lesions from dermoscopic images, achieving an accuracy of 80.79% on the HAM10000 dataset. The dataset consist saven classes. The study aims to demonstrate the potential of deep learning in accurately categorizing skin lesions, aiding in early detection and treatment of skin cancer. However, the focus on a specific dataset limits the generalizability of the proposed method to diverse clinical settings and populations. Despite this limitation, the research contributes to advancing the application of deep learning in dermatology, paving the way for future developments in computer-aided diagnosis and personalized healthcare.

An innovative method of classifying skin lesions using a pre- trained DenseNet201 architecture is presented in the publication ”Skin Lesion Classification Using a Pre-Trained

DenseNet201 Deep Neural Network" by Jasil, S.G.et al.,[3] With a training accuracy of 95% and a test accuracy of 77%, the suggested model shows promise in the early diagnosis of skin cancer. The research focuses on applying transfer learning and deep learning approaches to tackle the problem of correctly diagnosing seven types of skin lesions. To improve performance and generalizability, the study does concede that more validation using larger datasets and a comparison with alternative pre-trained architectures are necessary.

The author Wu, Y. et al.[4] Skin Lesion Classification based on Deep Convolutional Neural Networks uses transfer learning to improve the deep convolutional neural networks' ability to classify seven different types of skin lesions. With a validation accuracy of 86.69%, their enhanced ResNet50 model illustrated how transfer learning works to improve classification accuracy. In response to the rising prevalence of skin conditions, the research attempts to aid in the early diagnosis of skin cancer. The imbalanced sample sizes of the various types of skin lesions, however, represent a drawback of the study and could affect the predictive power of the model.

In this paper, a "Disease classification based on dermoscopic skin images using convolutional neural network in teledermatology system"is proposed by Purnama, I.K.E.et al.[5] The MNIST HAM10000 dataset was used to train and assess the suggested model. The model was evaluated by the authors using Cross Validation, and they were able to obtain accuracy of 91.5% for MobileNet v1 and 92.5% for Inception V3. The aim of this work is to offer a dependable and effective approach for the classification of skin diseases in teledermatology. The fact that this study is limited to seven skin illnesses means that it might not apply to other skin ailments.

The deep learning architectures MobileNet, VGG-16, and a custom model are used in "Skin lesion classification using deep learning architectures" Salian, A.C. et al.[6] proposed skin lesion categorization system. A few of the traits extracted, skin lesion prediction, and augmentation techniques included in the proposed system. The custom model was utilized by the writers to attain an accuracy of 80.61% with augmentation and 83.152% without it. In order to improve skin cancer diagnosis and therapy, this research aims to help with early skin lesion identification and classification. This research's drawback is the requirement for 20x magnification, which can only be accomplished with a certain tool and limits how the lesion can be analyzed.

A deep learning-based methodology for classifying dermoscopic pictures of skin lesions is presented in the publication "Skin Lesion Classification from Dermoscopic Images Using Deep Learning Techniques" by Quang, N.H., et al.,[7]. On the ISIC Archive dataset, the suggested method achieves a sensitivity of 78.66% and a precision of 79.74% by using the VGGNet convolutional neural network architecture with transfer learning. Enhancing medical imaging-based diagnosis systems and improving early melanoma detection are the goals of this research. Nevertheless, the limited size of the dataset limits the research and necessitates more validation on larger datasets.

A deep learning and transfer learning approach for skin cancer picture classification is presented in the publication "Transfer Learning Based Method for Two-Step Skin Cancer Images Clas- sification" by Mikołajczyk, A., et al.[8] The approach consists of two phases and is tested on the HAM10000 dataset. In the first step, the prediction model's accuracy is 85%, and in the second step, it is 75%. The project aims to enhance the detection of skin cancer through technological advancements. The study's modest sample size for the training period is its main drawback. In terms of classifying skin cancer images, the suggested technique performs well overall.

3. METHODOLOGY

The research's audit of skin lesion classification intends to improve diagnostic decision-making by using rigorous data preparation techniques such as augmentation, normalization, and feature engineering. The objective is to enhance skin lesion identification and classification using cutting-edge machine learning techniques, namely convolutional neural networks (CNNs), allowing medical practitioners to make more prompt and informed judgments. Figure 2 depicts the study workflow, which begins with data gathering from a classification dataset, namely HAM10000, and continues with pre-processing processes such as image ordering, resizing, data balance, and augmentation. The dataset is then partitioned into training and testing sets, and a model, such as a CNN or hybrid model, is trained using the pre-processed data. The trained model's performance evaluation covers met- rics such as accuracy, precision,recall, and F1-score, leading to a comprehensive assessment of the proposed methodology's success.

3.1 Dataset Description

The HAM10000 dataset, a cornerstone of dermatological research, is made up of a diversified group of 10,015 dermatoscopic pictures that have been rigorously selected to allow for a detailed exami- nation of skin diseases[9]. This dataset is divided into seven cate- gories: Melanoma, Melanocytic Nevi, Basal Cell Carcinoma, Be- nign Keratosis-Like Lesions, Actinic Keratoses, Dermatofibroma, and Vascular Lesions. Significantly, the dataset shows severe class imbalances, with Melanocytic Nevi appearing as the most com- mon group. This emphasizes the significance of developing strong techniques to overcome the issues posed by unequal class distribu- tions. This dataset's rich diversity and precise annotations not only provide a solid platform for the advancement of machine learning models, but also significantly add to our complete understanding of skin lesion categorization and detection tactics.

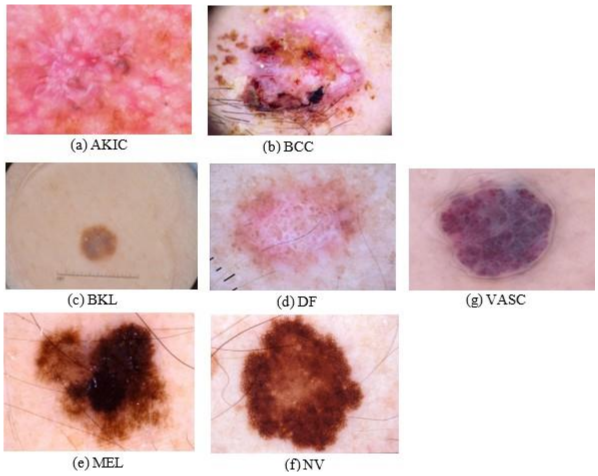

Fig. 1: Sample of Dataset

Table 1. : Number of Images for Each Class

Name of Classes	No of images
Actinic keratoses and intra epithelial carcinoma	327
Basal cell carcinoma	514
Benign keratosis-like lesions	1099
Dermatofibroma	115
Melanoma	1113
Melanocytic nevi	6705
Vascular lesions	142
Total	**10015**

3.2 PREPROCESSING

Image data preparation is one of the most important steps in prepar- ing raw images for in-depth analysis across our image collection. This first stage of preparation guarantees that the photos are cleaned and effectively transformed into an analysis-ready format.

Ordering the Dataset

Ordering the dataset by disease entails arranging the data accord- ing to a given criterion, which is usually based on disease cat- egories.This stage organizes the data for improved analysis and display[10].

Image Resizing

To ensure that the dermoscopy pictures met model input criteria, the dataset's original image resolution of 450 × 600 pixels was re- duced to a standardized size of 224 x 224 pixels. This scaling not only assures model compliance, but also improves computing per- formance during following operations, hence speeding the picture processing pipeline.

Data Balancing

A balancing approach was implemented to address the disparity in the HAM10000 dataset, which had 6705 pictures of melanocytic nevi but only 115 of Dermatofibroma.[15] This included undersam- pling the majority class and supplementing data for the minority class, resulting in an equitable distribution of 1400 photos for each. This balanced dataset seeks to reduce bias and improve model train- ing.

Smoothing Data

Smoothing data in image enhancement use Gaussian Blur to de- crease noise and emphasize important details. A unique sharpening kernel improves edges and fine details, replacing the original blurry images.

Table 2. : Data Split Ratios

Ratio	(70 : 30)%		(80 : 20)%		(90 : 10)%	
	Train	Test	Train	Test	Train	Test
HAM10000	7140	3060	8160	2040	9180	1020

Enhancement

Data augmentation is the process of increasing the quality or at- tributes of data to make it more helpful for analysis or model train- ing.

Dataset Split

In deep learning, dataset splitting is critical for model training and assessment. The split of training (70-90%) and testing (10-30%) sets promotes fair evaluation and prevents overfitting. Different ra- tios, such as 90:10, are used, as shown in Table 2, to assess model generality. The training set helps the model learn, whereas the test- ing set assesses its performance on previously unknown data. The 90:10 split, with its greater sample size, is especially important for reliable model evaluation and validation.

3.3 Model Building

In order to improve diagnostic decision-making, this research pa- per dives into the categorization of skin lesions, as described in the methodology section. Thorough data preparation, including aug- mentation and normalization methods, as well as feature engineer- ing to extract relevant information, is mandated by the method- ological approach. The major goal is to improve the precision and efficacy of skin lesion recognition and classification by lever- aging cutting-edge machine learning techniques, namely convo- lutional neural networks (CNNs). This, in turn, is aimed at as- sisting medical practitioners in making timely and informed de- cisions. An architectural depiction of the research workflow is provided in Fig.2. The initial phase involves the acquisition of data from a classification dataset, especially HAM10000. Pre- processing processes, such as picture ordering, image resizing, data balancing, smoothing data, and enhancement, are then applied to the classification dataset. Following pre-processing, the dataset is divided into training and testing sets. The model, which may be a CNN or a hybrid model with architectural choices such as Cus- tom CNN, DenseNet121, MobileNetV2, ResNet-50V2, VGG-16, ResNet50+DenseNet121+InceptionV3, is then trained using the pre-processed training data. The performance of the trained model is assessed using measures such as accuracy, precision, recall, F1- score, and AUC, resulting in a thorough assessment of the success of the proposed methodology.

3.4 DenseNet-121

DenseNet-121, a convolutional neural network architecture, is highly useful due to its parameter efficiency, enabled by dense connectivity, which reduces the number of parameters required. Its feature reuse mechanism facilitates better gradient flow, mitigating the vanishing gradient problem and enabling deeper networks. DenseNet-121 consistently achieves high accuracy in image classification tasks, surpassing other architectures.[11]

In research,the proposed DenseNet121's architectural struc- ture, supported by a sophisticated classifier, allows it to navigate the complicated environment of image categorization with ease. DenseNet121 organizes a neural ensemble, starting with the preloaded ImageNet weights, to extract rich feature hierarchies from input images. Then, a Global Average Pooling layer presents spatial complexities into a unified essence, establishing the frame- work for following layers[12]. A layer of dense connections arises, flaming with 1024 units triggered by ReLU activation, refining the collected characteristics into visible patterns. A 0.2-rate dropout layer interspersed inside this layer's structure protects the model from overfitting and ensures strong generalization. The final dense layer, embellished with softmax activation, culminates this orchestration by elegantly distilling the essence of subtle visual details into a harmonic ensemble of class probabilities. This architectural synthesis combines DenseNet121's capabilities with personalized classifier refinement, exemplifying innovation and efficacy in image classification.

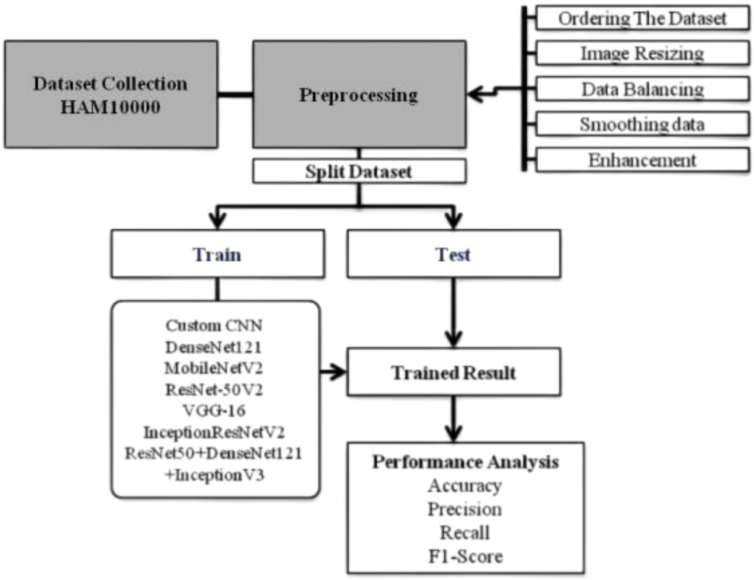

Fig. 2: The overview of the proposed Methodology

3.5 MobileNetV2

MobileNetV2 is a neural network model specifically developed for image categorization applications. This architecture combines convolutional power with classifier elegance, elegantly integrating feature extraction and classification. Anchored by MobileNetV2's excellent feature extraction, the network unleashes its power over a series of dense layers.[13]

In this proposed MobileNetV2's arcitecture the first layer, which has 512 units and ReLU activation, primes the network's discernment, followed by a prudent dropout layer that protects against overfitting with a delicately tuned rate of 0.3. As a result, a denser stratum arises, with 1024 units triggered by ReLU activation, enhancing the network's discriminative capabilities. A second dropout layer, with a nuanced rate of 0.2, improves the network's robustness to overfitting. The final dense layer, with 128 units and ReLU activation, refines the network's knowledge, condensing complicated image details into simple categories.[14] This architectural symphony arranges a harmonic combination of convolutional power and classifier delicacy, exemplifying innovation and efficacy in image classification.

3.6 ResNet50V2

ResNet50V2 uses the powerful ResNet50V2 as its base, enhancing its convolutional capabilities with an improved classifier. The fundamental model, supplied with ImageNet weights, functions as a feature extractor, expertly extracting complex image features. [16]

In the ResNet50V2 model, a Global Average Pooling layer follows, which consolidates spatial information into a compact representation. Following this, a dense layer arises, filled with 1024 units and driven by ReLU activation, which improves the network's discernment. A dropout layer, wisely weaved inside this stratum at a rate of 0.2, fortifies the model's robustness against overfitting. The last thick layer, embellished with softmax activa- tion, capably distills the essence of complicated visual specifics into a cohesive array of class probabilities. This architectural masterpiece combines the power of ResNet50V2 with bespoke classifier refinement, exemplifying innovation and efficacy in image classification.

3.7 VGG-16

VGG16 used as its foundation, reinforced by a unique classifier, to navigate the complex environment of image categorization with ease. VGG16, based on preloaded ImageNet weights, orchestrates an ensemble of convolutions to extract rich feature hierarchies from input images.[18]

In the proposed system, a Global Average Pooling layer con- denses spatial information into a compact representation, paving the way for the next thick stratum. With 1024 units activated by ReLU, this stratum refines the retrieved characteristics and shapes them into discriminative patterns. A dropout layer is interspersed across this stratum, its presence carefully calibrated at a rate of

0.2 to protect the network from overfitting. This architectural masterpiece combines the power of VGG16 with bespoke classifier refinement, exemplifying inventiveness and efficacy in the field of image classification.

3.8 InceptionResNetV2

InceptionResNetV2, blending Inception and ResNet architectures, excels in computer vision tasks like segmentation and object detection. Its deep residual connections ensure smooth gradi- ent flow, addressing vanishing gradient issues. With effective regularization, it prevents overfitting, enhancing generalization capabilities. This flexible solution offers pre-trained models and robust community support, empowering developers and re- searchers to achieve exceptional results in various applications.[20]

In this architectural, the imposing InceptionResNetV2 serves as the foundation, its weight filled with ImageNet's collected knowledge. The network flows through a symphony of levels, each carefully constructed for clarity and perfection. Beginning with the global average pooling layer, spatial subtleties are reduced to a compact representation, paving the way for the next layers.[21] A series of rich layers emerges, each resembling a brushstroke on the canvas of categorization. With 512 units blazing with ReLU activation, the network detects complicated patterns, aided by a well-placed dropout layer, and a rate of 0.3 that protects against overfitting. Following then, another stratum arises, with 1024 units pulsing with ReLU activation, helping to refine the network's knowledge. This stratum is protected by a dropout layer with a rate of 0.2, ensuring strong generalization. The final dense layer with 128 units, sparked by ReLU activation, culminates this symphony of layers. This architectural masterpiece, a monument to creativity and efficacy, combines the power of InceptionResNetV2 with proprietary classifier refinement.

4. RESULTS AND ANALYSIS

In this research, a wide range of models were used, including Mo- bileNetV2, ResNet-50V2, VGG-16, DenseNet-121, CNN, and a precisely created hybrid model. Multiple rounds of the CNN pro- gram with fine-tuned parameters were run to improve performance. The results were spectacular, with DenseNet-121 getting 91% ac- curacy and CNN and the hybrid model reaching 81%. Notably, the suggested model demonstrated remarkable illness prediction accu- racy, scoring 91%. These findings demonstrate the intricacy and usefulness of the technique for high-precision disease diagnosis.

4.1 COMPARING MODELS PERFORMANCE in DIFFERENT SPLIT RATIO

Table 3 depicts the dynamic performance development of four deep learning models—MobileNetV2, ResNet-50V2, VGG-16, and DenseNet-121—across different data splits (70:30, 80:20, 90:10) and epoch times. Notably, after 50 epochs, MobileNetV2 achieves 81% accuracy in the 90:10 split, whereas ResNet-50V2 steadily increases to an amazing 89%. VGG-16 achieves a con- sistent 78% accuracy across all splits. DenseNet-121 stands out, with a significant improvement from 84% to an impressive 91% accuracy in the 90:10 split after 50 epochs. The inclusion of Incep- tionResNetV2 improves the comparison landscape by adding 77% accuracy in the 70:30 split, 80% in the 80:20 split, and 85% in the 90:10 split after 50 epochs. This concise summary offers useful in- sights for model selection and training configuration optimization based on unique dataset features.

Table 3. : Model performances in Different Split Ratio

Model	**(70 : 30)%**		**(80 : 20)%**		**(90 : 10)%**	
	30	**50**	**30**	**50**	**30**	**50**
MobileNetV2	75.62	79.57	75.68	82.79	79.11	81.66
RasNet-50V2	83.23	84.34	85.24	89.06	88.23	89.21
VGG-16	78.36	80.16	80.09	80.44	80.23	80.39
InceptionResNetV2	77.25	80.35	76.22	80.34	80.39	85.19
DenseNet-121	84.08	85.88	85.63	89.06	87.64	**91.17**

4.2 COMPARING CNN and HYBRID PERFORMANCES in DIFFERENT SPLIT RATIO

The table 4 shows that CNN and ResNet50+DenseNet121+InceptionV3 models exhibit signifi- cant performance patterns over a range of data split ratios and epochs. After 150 epochs, CNN exhibits constant accuracy im- provement, with split ratios of 78%, 81%, and 77%. In contrast, the hybrid model exhibits dynamic accuracy patterns, obtaining 79% at 30 epochs, peaking at 81% at 100 epochs, and retaining 79% at 50 epochs. The comprehensive research informs model selection based on unique dataset requirements.

Table 4: Model performances in Different Split Ratio

Model	**70% - 30%**		**80% - 20%**		**90% - 10%**	
	Epoch	**Accuracy**	**Epoch**	**Accuracy**	**Epoch**	**Accuracy**
CNN	150	78.45	150	81.22	150	77.31
ResNet50+ DenseNet121+ InceptionV3	30	80.68	100	81.91	50	79.80

4.3 EXPERIMENTS on CNN

Table 5 illustrates the results of many trials on a CNN model, in- cluding layer configurations, batch size, epochs, dropout rates, and learning rates. Key findings include the effect of layer connections, dropout techniques, and learning rate schedules on accuracy per- centages. The simplified style allows for quick comparison of trial outcomes, assisting in the rapid discovery of optimal parameter combinations for the CNN model.

Table 5. : Experiment Results with Different Parameters

Ex No.	**Layers**	**Batch Size**	**Epoch**	**Dropout**	**Learning Rate**	**Accuracy**
01	03	32	50	-	-	70.63
02	04	32	50	0.5	0.0001	58.28
03	04	32	50	0.5	Schedule	58.52
04	03	32	100	0.3	0.01	63.33
05	03	32	20	0.5	0.001	69.46
06	04	32	50	0.5, 0.3	0.0001	72.64
07	04	32	150	0.7	0.0001	74.90
08	04	32	100	0.7	0.0001	75.68
09	04	32	250	0.25, 0.5	0.001	76.96
10	04	32	150	0.5, 0.3	0.0001	**81.22**

4.4 ANALYSIS of MODEL with PRECISION RECALL F1-SCORE in SPLIT RATIO

The table 6 displays the Analysis deep learning model performance across various training-test data splits (80% - 20% and 90% - 10%). DenseNet121 consistently achieves high accuracy, recall, and F1- Score across both splits, demonstrating its durability. InceptionV3 performs well, particularly in the 90% - 10% split, demonstrat- ing its applicability for bigger datasets. However, other models, such as CNN, show slight performance changes across splits, in- dicating sensitivity to dataset size. Furthermore, the Hybrid model ResNet50 + DenseNet121 + InceptionV3 produces encouraging re- sults, demonstrating the potential benefits of model ensemble ap- proaches in improving classification performance across a variety of data splits.

4.5 Training and validation accuracy Curves

(a) MobileNetV2

In the above Fig 3 shows that, after 50 epochs, the model's training accuracy of 0.9214 demonstrated how well it could learn from the dataset. Good generalization to new, unseen data is indicated by the validation accuracy, which was 0.8167. The model's ability to learn and perform effectively on the given task is suggested by the steady improvement in training and validation accuracy throughout epochs.

Table 6. : Model performances in Different Split Ratio

Model	80% - 20%			90% - 10%		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
MobileNetV2	84	83	83	84	82	82
ResNet50V2	89	89	89	90	89	89
InceptionV3	88	87	87	90	89	89
VGG16	82	80	80	82	80	80
InceptionResNet V2	82	80	80	85	85	85
DenseNet121	90	89	89	92	91	91
CNN	82	81	81	79	77	77
ResNet50 + DenseNet121 + InceptionV3	83	82	82	82	80	80

(b) ResNet50V2

Fig 4 shows the result, after training for more than 50 epochs, the model's accuracy was impressive. Significant learning from the dataset was indicated by the final training accuracy, which attained an astounding 0.9864. Concurrently, the validation accu- racy reached a high value of 0.8922, demonstrating the model's successful generalization to previously encountered data. The increasing trend in validation accuracy and accuracy over the course of the epochs highlights the model's ongoing learning and improvement.

(c) VGG-16

After 50 epochs, the model reached a training accuracy of 0.8809, as shown in Fig 5, indicating that it has learned the training set effectively. Good generalization to new, unseen data is indicated by the validation accuracy of 0.7856. The model's capacity to pick up new skills and perform effectively on the assigned task is suggested by the steady improvement in training and validation accuracy over the course of the epochs.

(d) InceptionResNetV2

Fig 6 illustrates that the model successfully learned from the dataset, achieving a training accuracy of 0.8956 after 50 epochs. Strong generalization to fresh, untested data is indicated by the validation accuracy of 0.8520. The model's capacity to pick up new skills and perform effectively on the assigned task is suggested by the steady improvement in training and validation accuracy throughout epochs.

(e) DenseNet121

Fig 7 shows that the model consistently improved in accuracy over the course of 50 epochs of training. The ultimate training accuracy of 0.9773 demonstrated strong learning from the dataset. Simultaneously, the validation accuracy hit a high of 0.9118, highlighting the model's effective generalization to new data. The increasing pattern in validation accuracy and accuracy over epochs highlights the model's ongoing improvement and learning process. **(f)CNN**

The accuracy curve for CNN is shown in Fig 8. Dropout and BatchNormalization technique is applied on fully connected layers in this model. This model is same as CNN 4th experiment, to improve the result epoch number are increased in range of 150. By the 150th epoch, it has improved to a commendable 0.9814 accuracy from an initial 0.2268 accuracy. In a similar vein, the validation accuracy increases from 0.3873 to 0.8088, suggesting that the model has good data generalization capabilities. **(g)ResNet50 + DenseNet121 + InceptionV3**

The accuracy curve for ResNet50+DenseNet121+InceptionV3 is shown in Fig 9[23]. In this model 1 connected layer is used. By adding learning rate and dropout is present in connected layer which helps prevent overfitting and giving a better accuracy. By the 100th epoch, it has improved to a commendable 0.9472 accuracy from an initial 0.5034 accuracy. In a similar vein, the validation accuracy increases from 0.6446 to 0.8191, suggesting that the model has good data generalization capabilities

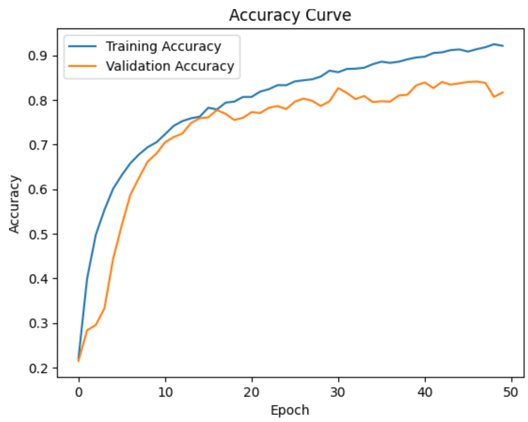

Fig. 3: Training and Validation Accuracy of MobileNetV2

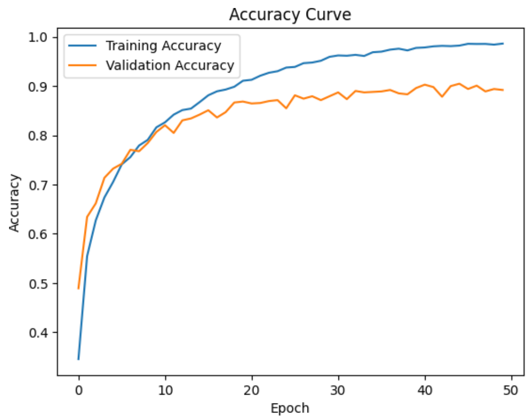

Fig. 4: Training and Validation Accuracy of ResNet50V2

4.6 Training and validation loass Curves

(a) MobileNetV2

Fig 10 is a graphical representation of both training loss and validation loss show a progressive decline during the training process. A reduction in the training loss from 1.9058 to 0.2057 signifies that the training data has been well learned. Concurrently, the validation loss drops from 1.9236 to 0.5320, indicating that the model performs well in terms of generalizing to previously unseen validation data.

(b) ResNet50V2

The Fig 11, the training loss continuously drops from 1.7033 to

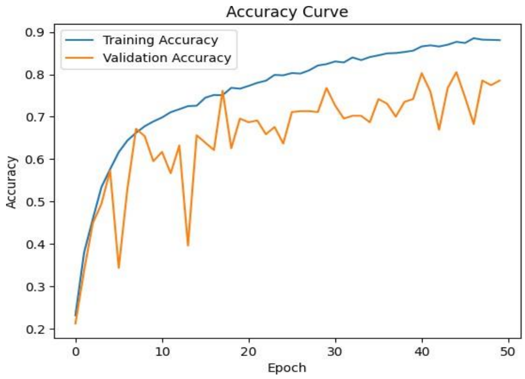

Fig. 5: Training and Validation Accuracy of VGG16

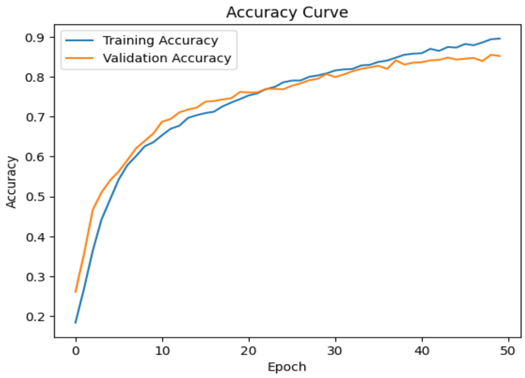

Fig. 6: Training and Validation Accuracy of InceptionResNetV2

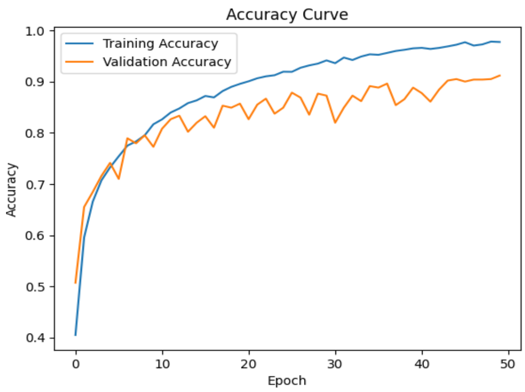

Fig. 7: Training and Validation Accuracy of DenseNet121

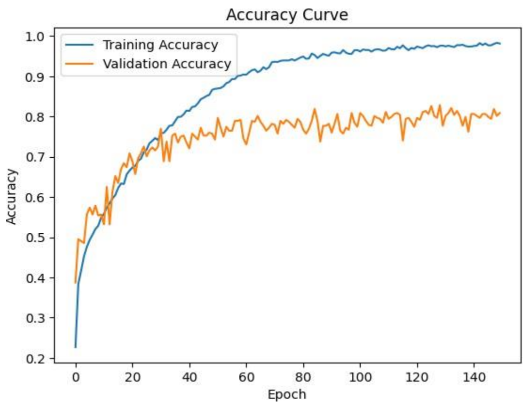

Fig. 8: Training and Validation Accuracy of CNN

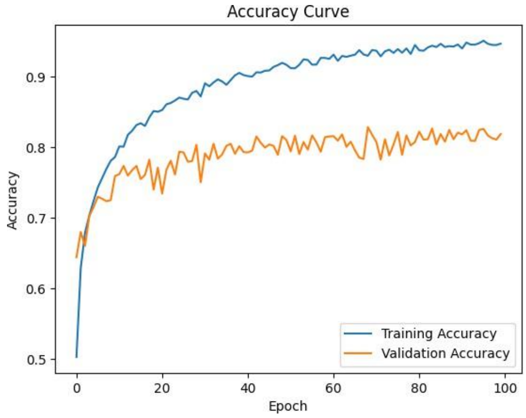

Fig. 9: Training and Validation Accuracy of ResNet50+ DenseNet121+InceptionV3

0.0444 throughout the course of the 50 epochs, demonstrating successful learning on the training dataset. Concurrently, there is a downward trend in the validation loss (validation loss), which goes from 1.4323 to 0.3609. Robust performance on both training and validation sets is seen by the declining loss and validation loss numbers.

(c) VGG-16

According to Fig 12, Both training loss and validation loss steadily decline throughout the course of the 50 epochs. Validation loss falls from 1.8784 to 0.6346, whereas training loss begins at 1.8585 and ends at 0.3257. The final low values of both training and validation loss suggest that the model effectively captures patterns in the training data and performs well on unseen data, demonstrating a successful training process.

(d) InceptionResNetV2

According to Fig 13, both training loss and validation loss contin- uously decline throughout the course of the 50 epochs, suggesting efficient learning. Validation loss drops from 1.8888 to 0.4258, while training loss begins at 1.9347 and continues to 0.2893. The model has successfully trained when the final low values of training and validation loss show that it has recognized the patterns in the training data and can generalize to new, unknown data. **(e)DenseNet121**

Fig 14 illustrates that over 50 epochs, the training loss continuously drops from 1.5768 to 0.0735, suggesting better fitting to the train- ing set. Concurrently, there is a downward trend in the validation loss, which goes from 1.2862 to 0.3122. Strong performance on both training and validation sets, as well as successful model training, are shown by the dropping loss and validation loss values. **(f)CNN**

In Fig 15 showing the loss curve of CNN model architecture are well suited for the complexity of the classification of skin lesion. This curve shown a little bit of fluctuation. The validation loss decreases from 1.6872 to 0.9895.

(e) ResNet50 + DenseNet121 + InceptionV3

In Fig 16 showing the loss curve of ResNet50+DenseNet121+InceptionV3 model architecture is well suited for the complexity of the classification of skin lesion. For adding dropout in connected layer helps prevent overfitting and decreasing the loss rates 1.9670 to 0.4449.

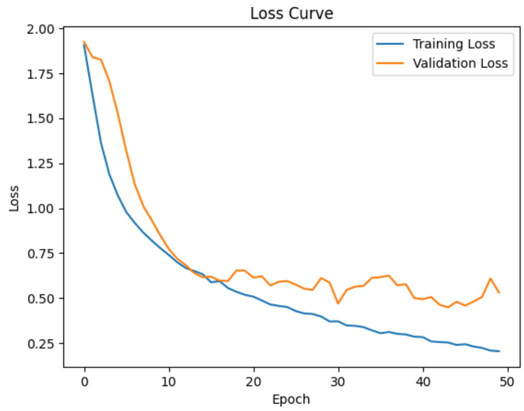

Fig. 10: Training and Validation Loss of MobileNetV2

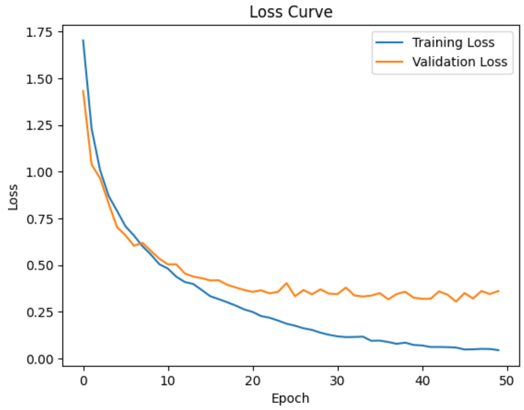

Fig. 11: Training and Validation Loss of ResNet50V2

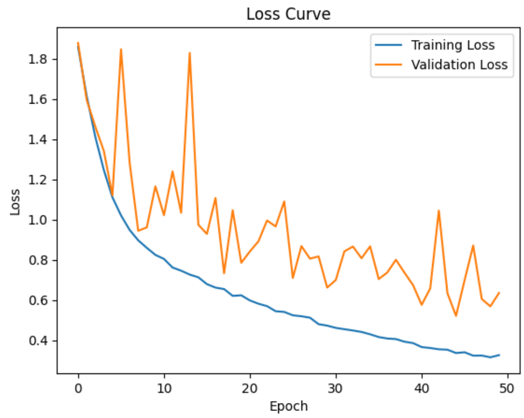

Fig. 12: Training and Validation Loss of VGG16

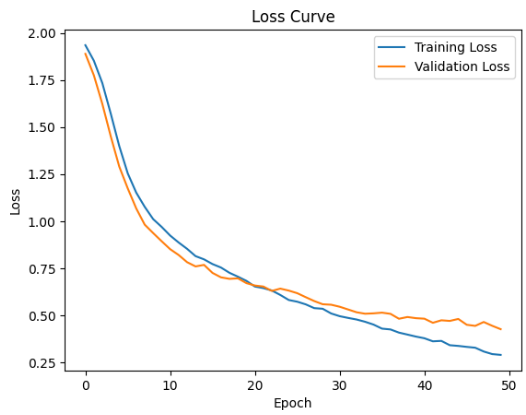

Fig. 13: Training and Validation Loss of InceptionResNetV2

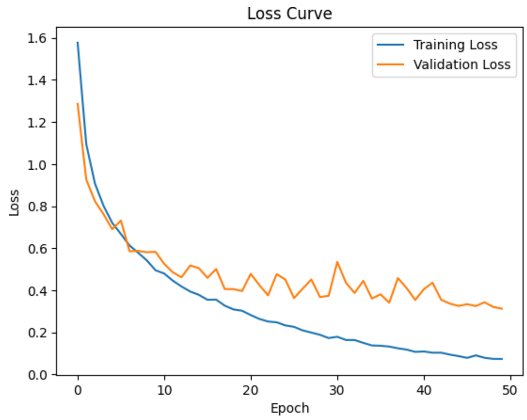

Fig. 14: Training and Validation Loss of DenseNet121

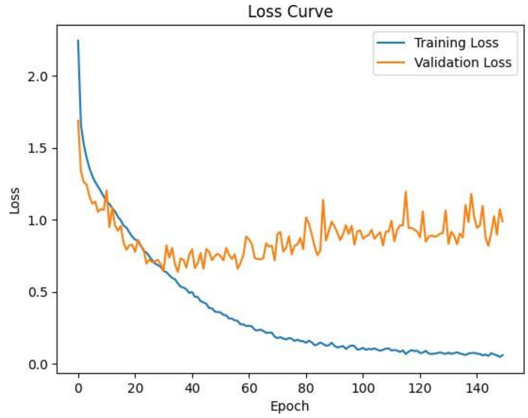

Fig. 15: Training and Validation Loss of CNN

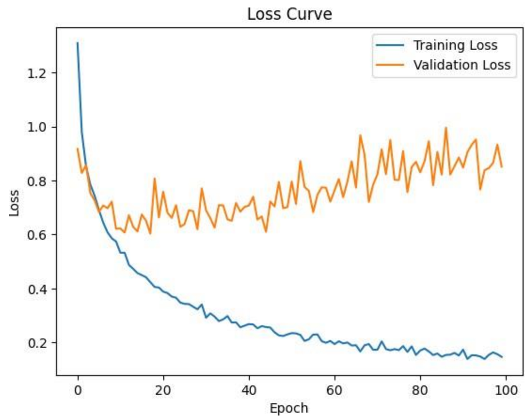

Fig. 16: Training and Validation Loss of ResNet50+ DenseNet121+InceptionV3

4.7 CONFUSION METRICS

Confusion metrics are a simple table used in machine learning to assess categorization model performance. It summarizes model predictions by comparing them to actual outcomes, which include true positives, true negatives, false positives, and false negatives. These measures are critical for generating performance indicators including as precision, recall, accuracy, and the F1 score, which provide a thorough assessment of a model's classification skills.

(a) MobileNetV2

The confusion matrix in Fig 17 high diagonal values, which correspond to accurate predictions, demonstrate the effectiveness of the classification model. It's interesting to observe that it can be difficult to discern between these pairs because the model shows some misunderstanding across a few classes. Even with these little misclassification errors, the accuracy is still quite high overall. The model performs well overall, indicating a reliable classifier with certain areas that still require improvement

(b) ResNet50V2

In the given confusion matrix of Fig 18, indicates that the classification model is operating well, as seen by the high diagonal values that correspond to accurate predictions. Interestingly, the model shows some confusion between few classes that it would be difficult to tell these couples apart. Even with these small errors in classification, the total accuracy is still excellent. The model performs well overall, indicating a dependable classifier with certain areas in need of improvement.

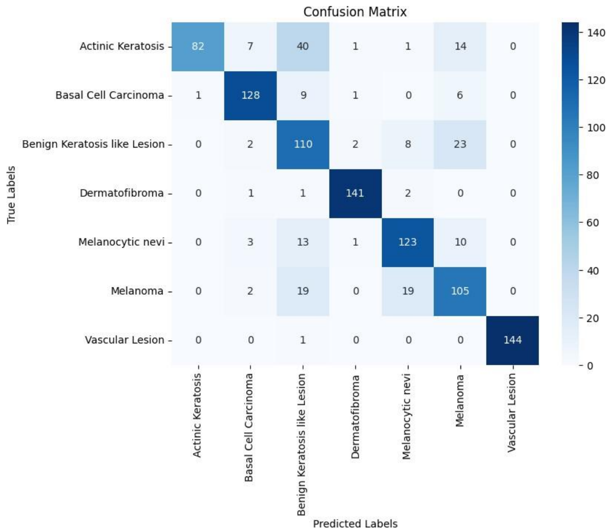

Fig. 17: Confusion Matrix of MobileNetV2

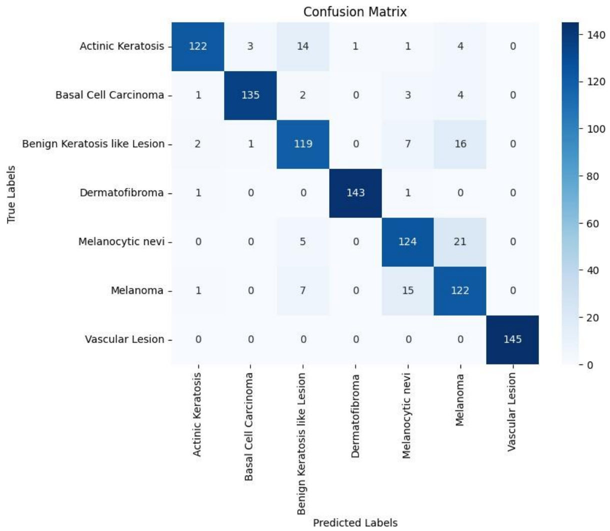

Fig. 18: Confusion Matrix of ResNet50V2

(c) VGG-16

The efficacy of the classification model can be seen in Fig 19's high diagonal values of the confusion matrix, which exactly meet predictions. It's interesting to note that the model exhibits a lot of uncertainty in a few classes, which makes selecting between these combinations challenging. Despite these minor classification's mistakes, the accuracy is still rather high overall. Overall, the model's performance is good, suggesting that while certain aspects still require work, it is a trustworthy classifier

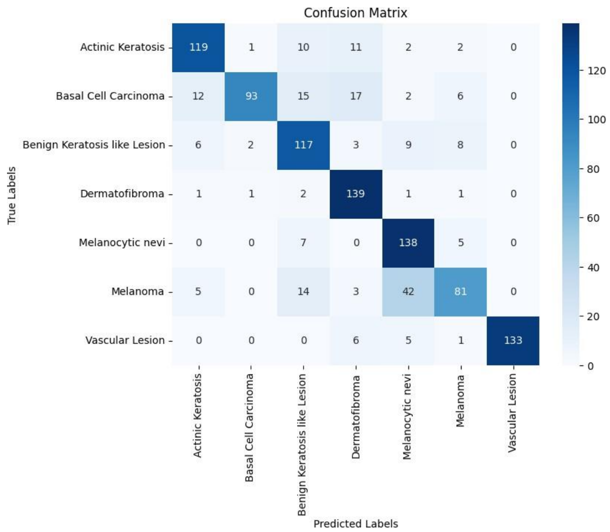

Fig. 19: Confusion Matrix of VGG16

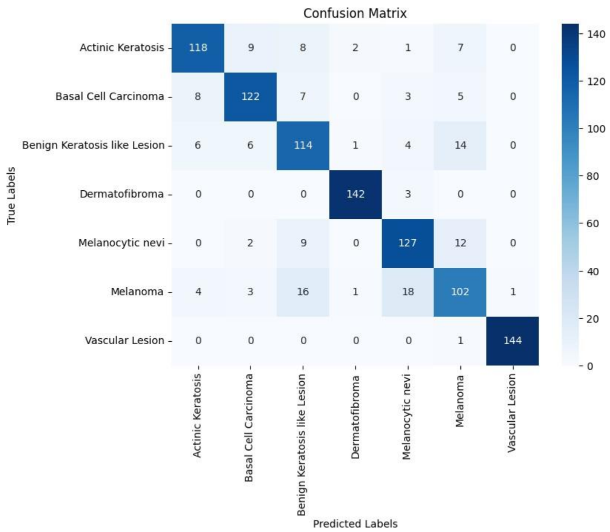

Fig. 20: Confusion Matrix of InceptionResNetV2

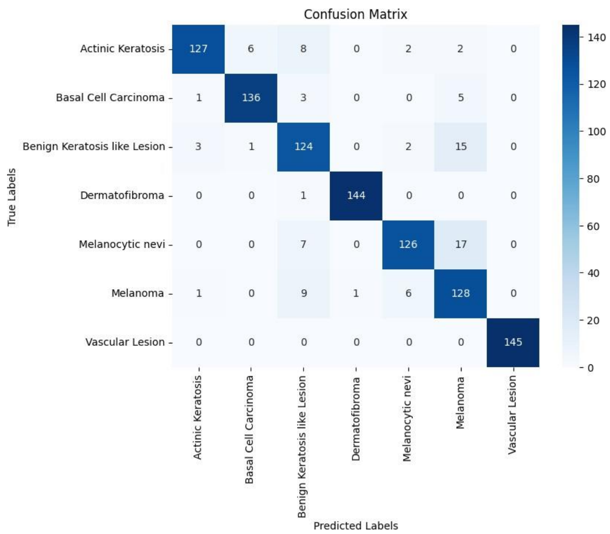

Fig. 21: Confusion Matrix of DenseNet121

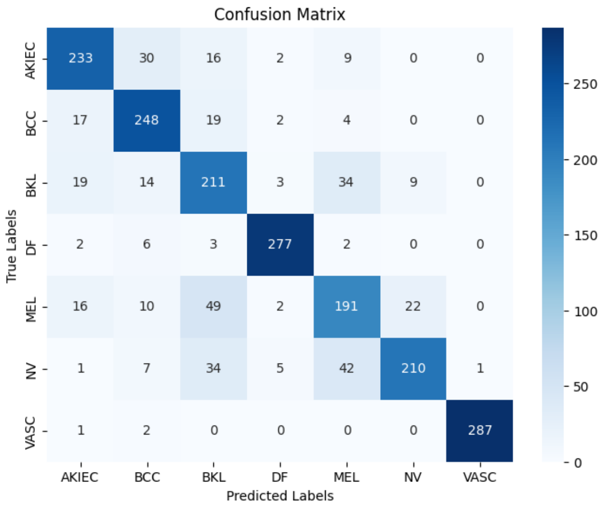

Fig. 22: Confusion Matrix of CNN

(d) InceptionResNetV2
The efficacy of the classification model is seen by the high diagonal values of the confusion matrix in Fig 20, which match precise predictions. It's interesting to note that the model indicates considerable confusion across a few classes, making it challenging to distinguish between these couples. Despite these minor misclas- sification mistakes, the total accuracy is still rather good. Overall, the model's performance is good, suggesting that it is a trustworthy classifier with some areas still in need of development.

(e) DenseNet121
Fig 21 displays confusion metrics of the model of DenseNet121. Some classes are accurately identified by the model, even with even little misclassifications. Although there are certain areas that could use development, overall, the matrix shows a strong model with excellent overall accuracy.

(f) CNN
Fig 21 provides a more detailed study of the model's performance and presents a summary of the classification problem's outcomes. It also shows the confusion metrics of CNN 4th experiment. Although it has some inaccurate predictions, the model has a rea- sonable aptitude for both positive and negative scenario prediction. **(g)ResNet50 + DenseNet121 + InceptionV3**

Fig 23 provides a more detailed study of the model's per- formance and presents a summary of the classification prob- lem's outcomes. It also shows the confusion metrics of ResNet50+DenseNet121+InceptionV3. Although it has some inaccurate predictions, the model has a reasonable aptitude for both positive and negative scenario prediction.

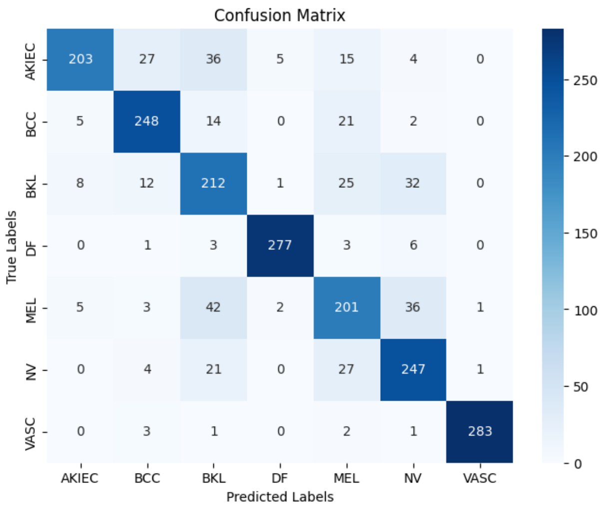

Fig. 23: Confusion Matrix of ResNet50+ DenseNet121+ InceptionV3

5. COMPARISON WITH PREVIOUS WORKS

The table 6 compares the accuracy percentages of machine learning models published by Ahmet Furkan SO¨ NMEZ, Neeshma A, and "Our Proposed". Notably, "Our Proposed" model performs well with 89% in ResNet50V2, 91% in DenseNet121, and 82% in Mo- bileNetV2. While others produce competitive outcomes, "Our Pro- posed" methodology shows potential across several architectures. Considerations like dataset qualities and training techniques are critical for a thorough evaluation.

Table 7. : Comparison of Model Performances

Models	**Ahmet Furkan SO¨ NMEZ[2**	**Neeshma A.[1]**	**Our Pro- posed**
MobileNetV2	80	-	82
InceptionResNetV2	76.43	-	85
VGG16	57	-	80
CNN	-	-	81

ResNet50+ DenseNet121+ InceptionV3	-	-	81
ResNet50V2	62	-	89
DenseNet121	-	82.1	**91**

6. CONCLUSION

The integration of deep learning (DL) in skin lesion prediction represents a significant advancement in dermatological diagnos- tics, promising enhanced accessibility, continuous learning, speed, and accuracy. The innovative approach employs DenseNet169, DenseNet121, ResNet101, a Hybrid model, and CNNs to of- fer precise predictions and encourage proactive self-care. No- tably, ResNet101 achieved 91% accuracy, while DenseNet169 and DenseNet121 attained 90% accuracy in respective data splits. Addi- tionally, a hybrid model demonstrated an impressive 88% accuracy, overcoming dataset imbalances using strategic evaluations. The cu- rated dataset not only aids in prediction but also highlights preven- tive methods and disease details. CNNs contributed significantly with an 81% accuracy, underscoring their importance in predictive modeling. Future work involves the development of AI-based apps and the exploration of hybrid models to further improve skin le- sion classification. Increasing dataset sizes and investigating real- world datasets are crucial for enhancing CNNs' resilience and per- formance. Studying various hybrid model configurations offers po- tential avenues for advancing skin lesion classification accuracy.

7. REFERENCES

[1] Neeshma, A. and Nair, C.S., 2022, August. Multiclass Skin Lesion Classification Using Densenet. In 2022 Third Interna- tional Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT) (pp. 506-510). IEEE.

[2] SO¨ NMEZ, A.F., C¸ AKAR, S., CEREZC˙I, F., KOTAN,

[3] M., DEL˙IBAS¸ OG˘ LU, ˙I. and Gu¨lu¨zar, C¸ . ˙I.T., 2023. Deep Learning-Based Classification of Dermoscopic Images for Skin Lesions. Sakarya University Journal of Computer and Information Sciences, 6(2), pp.114-122.

[4] Jasil, S.G. and Ulagamuthalvi, V., 2021, May. Skin lesion classification using pre-trained DenseNet201 deep neural net- work. In 2021 3rd international conference on signal process- ing and communication (ICPSC) (pp. 393-396). IEEE.

[5] Wu, Y., Lariba, A.C., Chen, H. and Zhao, H., 2022, July. Skin Lesion Classification based on Deep Convolutional Neu- ral Network. In 2022 IEEE 4th International Conference on Power, Intelligent Computing and Systems (ICPICS) (pp. 376-380). IEEE.

[6] Purnama, I.K.E., Hernanda, A.K., Ratna, A.A.P., Nurtanio, I., Hidayati, A.N., Purnomo, M.H., Nugroho, S.M.S. and Rach- madi, R.F., 2019, November. Disease classification based on dermoscopic skin images using convolutional neural network in teledermatology system. In 2019 international conference on computer engineering, network, and intelligent multimedia (CENIM) (pp. 1-5). IEEE.

[7] Salian, A.C., Vaze, S., Singh, P., Shaikh, G.N., Chapaneri, S. and Jayaswal, D., 2020, April. Skin lesion classification using deep learning architectures. In 2020 3rd International confer- ence on communication system, computing and IT applica- tions (CSCITA) (pp. 168-173). IEEE.

[8] Quang, N.H., 2017, November. Automatic skin lesion anal- ysis towards melanoma detection. In 2017 21st Asia Pacific symposium on intelligent and evolutionary systems (IES) (pp. 106-111). IEEE.

[9] Mikołajczyk, A. and Grochowski, M., 2018, May. Data aug- mentation for improving deep learning in image classification problem. In 2018 international interdisciplinary PhD work- shop (IIPhDW) (pp. 117-122). IEEE.

[10] "Skin Cancer MNIST: HAM10000," www.kaggle.com. https://www.kaggle.com/datasets/kmader/skin-cancer-mnist- ham10000

[11] Shetty, B., Fernandes, R., Rodrigues, A.P., Chengoden, R., Bhattacharya, S. and Lakshmanna, K., 2022. Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. Scientific Reports, 12(1), p.18134.

[12] Mahbod, A., Schaefer, G., Wang, C., Ecker, R., Dorffner, G. and Ellinger, I., 2021, January. Investigating and exploiting image resolution for transfer learning-based skin lesion clas- sification. In 2020 25th international conference on pattern recognition (ICPR) (pp. 4047-4053). IEEE.

[13] Zhang, Y. and 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.

[14] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C., 2018. Mobilenetv2: Inverted residuals and linear bottle- necks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).

[15] Ogundokun, R.O., Li, A., Babatunde, R.S., Umezuruike, C., Sadiku, P.O., Abdulahi, A.T. and Babatunde, A.N., 2023. En- hancing skin cancer detection and classification in dermo- scopic images through concatenated MobileNetV2 and xcep- tion models. Bioengineering, 10(8), p.979.

[16] Arya, M.S., Prabahavathy, P. and Ahamed, S., 2022. Skin le- sion classification and Prediction by Data Augmentation in HAM10000 and ISIC 2019 dataset.

[17] Kondaveeti, H.K. and Edupuganti, P., 2020, December. Skin cancer classification using transfer learning. In 2020 IEEE In- ternational Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI) (pp. 1-4). IEEE.

[18] Shah, V., Autee, P. and Sonawane, P., 2020, December. Detec- tion of melanoma from skin lesion images using deep learning techniques. In 2020 International Conference on Data Science and Engineering (ICDSE) (pp. 1-8). IEEE.

[19] Quang, N.H., 2017, November. Automatic skin lesion anal- ysis towards melanoma detection. In 2017 21st Asia Pacific symposium on intelligent and evolutionary systems (IES) (pp. 106-111). IEEE.

[20] Salian, A.C., Vaze, S., Singh, P., Shaikh, G.N., Chapaneri, S. and Jayaswal, D., 2020, April. Skin lesion classification using deep learning architectures. In 2020 3rd International

confer- ence on communication system, computing and IT applica- tions (CSCITA) (pp. 168-173). IEEE.

[21] Ahmed, S.A.A., Yanikoğlu, B., Göksu, Ö. and Aptoula, E., 2020, October. Skin lesion classification with deep CNN en- sembles. In 2020 28th Signal Processing and Communica- tions Applications Conference (SIU) (pp. 1-4). IEEE.

[22] Nguyen, V.D., Bui, N.D. and Do, H.K., 2022. Skin lesion classification on imbalanced data using deep learning with soft attention. Sensors, 22(19), p.7530.

[23] Sachdev, J., Shekhar, S. and Indu, S., 2018, April. Melanoma screening using deep neural networks. In 2018 3rd Interna- tional Conference for Convergence in Technology (I2CT) (pp. 1-5). IEEE.

[24] Mahbod, A., Schaefer, G., Wang, C., Ecker, R. and Ellinge, I., 2019, May. Skin lesion classification using hybrid deep neu- ral networks. In ICASSP 2019-2019 IEEE international con- ference on acoustics, speech and signal processing (ICASSP) (pp. 1229-1233). IEEE.

[25] Lopez, A.R., Giro-i-Nieto, X., Burdick, J. and Marques, O., 2017, February. Skin lesion classification from dermo- scopic images using deep learning techniques. In 2017 13th IASTED international conference on biomedical engineering (BioMed) (pp. 49-54). IEEE.

[26] Moldovan, D., 2019, November. Transfer learning based method for two-step skin cancer images classification. In 2019 E-Health and Bioengineering Conference (EHB) (pp. 1- 4). IEEE.

[27] Jain, S., Singhania, U., Tripathy, B., Nasr, E.A., Aboudaif, M.K. and Kamrani, A.K., 2021. Deep learning-based trans- fer learning for classification of skin cancer. Sensors, 21(23), p.8142