Melanoma Skin Cancer Detection Using Ensemble Deep Learning Model

Md Mahbubur Rahman
Department of Computer Science
(ACIT-Applied AI)
Oslo Metropolitan University
Oslo, Norway
S371512@oslomet.no

Quang Dung Martin Phan
Department of Computer Science
(ACIT-Applied AI)
Oslo Metropolitan University
Oslo, Norway
S333747@oslomet.no

Mohammad Azizul Islam Yasin
Department of Computer Science
(ACIT-Applied AI)
Oslo Metropolitan University
Oslo, Norway
S371138@oslomet.no

***Abstract*— Skin cancer like melanoma is a potentially fatal disease that, if diagnosed and treated at an early stage, offers the best chance of a cure. Melanoma is difficult to diagnose early since its symptoms are not easily noticeable to the untrained eye. The primary motivation for this study is to develop a faster and more accurate method for detecting melanoma skin cancer in its earliest stages. In the past, machine learning strategies have been applied to the identification of skin cancer based on protein sequences and other imaging modalities. The problem with machine learning methods is that they necessitate human-engineered features, which is an extremely tedious and time-consuming process. Deep learning was able to partially solve this problem by automatically generating feature extractions. By combining the knowledge of multiple individual models, the ensemble learning method can make more accurate decisions. Consequently, the accuracy of melanoma detection can be improved by combining the decisions of multiple learners.**

In this study, we present an ensemble deep learning model based on transfer learning technique for classifying skin lesions as melanoma or non-melanoma utilizing the VGG19, ResNet50, and MobileNetV2 pre-trained models. The suggested model performed very well on the ISIC 2020 dataset, with an accuracy of 99.4%, proving its use in the identification of melanoma at an early stage. The experimental results show that the proposed ensemble deep learning model is more accurate and requires less computing than the state-of-the-art deep learning techniques.

***Keywords*— Machine learning, deep learning, pre-trained model, ensemble learning, melanoma detection.**

I. Introduction

In the current decade, skin cancer has been one of the most rapidly growing cancers. The skin is the largest organ in the body, therefore it makes sense that it is also the most commonly affected by cancer[1]. The skin is the body's first line of defense against the elements, including UV rays, heat, damage, and microbial or viral diseases. The skin stores energy in the form of fat and water, and it also helps the body regulate its temperature. There are eight different types of skin cancer: melanoma (MEL), melanocytic nevi (NV), basal cell carcinoma (BCC), benign keratosis lesions (BKL), actinic keratosis (AK), dermatofibroma (DF), squamous cell carcinoma (SCC), and vascular lesions (VASC) [3]. MEL is the deadliest form of cancer because it quickly invades other organs. Cells in the body called melanocytes are responsible for their development.

Figure 1: Benign Image

Figure 2: Malignant Image

Figures 1 and 2 depict dermoscopy image samples of benign and malignant conditions, respectively.

Melanin, the skin's pigment, begins to develop cytes, which is the first step in the development of melanoma. It can penetrate deep into the skin, enter the bloodstream, and spread to other parts of the body. Melanoma skin cancer has a high propensity to spread to other organs and tissues if it is not detected and treated early. The number of people diagnosed with melanoma rises each year. It has been estimated that there will be 97,610 new cases of melanoma identified in the United States and 7990 deaths from the disease in 2023 by the Melanoma Foundation, a renowned cancer research center.

Patients with melanoma have a much better chance of survival if the cancer is diagnosed and treated early. New methods and techniques are being developed to complement self-assessments in the quest to detect melanoma at an early stage, which is crucial for patients' chances of survival.

Medical professionals employ diverse methods to identify instances of skin cancer. An experienced dermatologist will do a primary visual checkup, then dermoscopy, and ultimately a biopsy to determine the nature of a suspicious growth. It may take a while, and in the meantime, the person may progress to the next level.

Due to the complexity and subjectivity of human interpretation, computerized analysis of dermoscopy images has become an essential area of study [5] to reduce diagnostic errors. To overcome this difficulty, researchers have dedicated a great deal of time and energy to learning how to classify tumours as malignant or benign using computer image analysis techniques. These systems perform segmentation, identification, and classification of melanoma using a variety of methods by combining image processing, computer vision, and machine learning [7].

The main problems with this research are the insufficient data and the lack of diversity in skin cancer progression. The traditional method of categorizing is fallible because it is based on human observation. Dermoscopy is a noninvasive imaging technique utilized to improve cancer diagnosis. Dermoscopy is a method for enhancing the visibility of a skin mark by capturing a magnified and illuminated image of the affected area of skin.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Several machines learning methods, including KNN, SVM, clustering, regression, and deep learning, have been developed and evaluated by researchers for both benign and malignant skin cancers [8, 9]. To better identify and categorize cancers, researchers have focused on a small range of textural, visual, and color traits.

Deep learning is the most prevalent and robust classification technique based on execution time, system complexity, and system performance [9]. Traditional classification methods only allowed for the transformation of data from its original format into a solution that included predicted and classifier end outcomes [10, 12]. The model is enhanced by the deep learning approach, which creates a fully automated classification and prediction system. To improve performance and obtain highly accurate results, researchers are turning to the development of hybrid systems that combine deep learning techniques [13].

Unfortunately, deep learning models' effectiveness is severely constrained by the scarcity of relevant medical data. Since there are insufficient labeled medical images, researchers frequently employ transfer learning strategies. Transfer learning is the process of retaining information learned while solving one problem and transferring it to a different but related problem. In addition, when it comes to making decisions on delicate matters like cancer detection, the abilities of learners are severely restricted. Individual learners' decisions can be combined to overcome this issue. The ensemble models are made using different types of learners with different properties to catch the differences in shape, structure, and texture in the skin cancer images. It is anticipated that the collective decision will be more precise than that of the individual models. Combining the decisions of discrete learners can improve the accuracy of skin cancer detection.

In the presented work, an ensemble of deep learners utilizing VGG19, ResNet50 and MobileNetV2 has been developed. The findings show that the suggested ensemble model outperformed the individual deep learners in terms of classification accuracy.

The following parts of the paper are organized as follows: Section II presents the related work, Section III describes the proposed methodology, Section IV presents the results and discussion, and Section V reviews the conclusion and future work.

II. RELATED WORKS

Many studies on the detection and diagnosis of melanoma skin cancer have been conducted in the last decade. The numerous datasets are provided for the research community. Researchers have applied strategies based on splitting, merging, clustering, and classification to the identification and treatment of skin cancer. Each approach has its own set of limitations and advancements. In general, the visual inspection of potentially malignant lesions carried out using an optical dermatoscope is a challenging task and requires a specialist dermatologist. For melanoma, a particularly aggressive type of skin cancer, only about 60–90% of malignant cancers are identified based on visual inspection, and accuracy varies markedly depending on the experience of the dermatologist.

To address this issue, a tremendous amount of research has been conducted on skin cancer detection utilizing machine learning and deep learning methodologies. Melanoma detection is accomplished through machine learning techniques that involve the extraction of manual features from dermoscopy images. In their publication, Waheed et al. proposed a machine learning methodology for the identification of melanoma in dermoscopic images [10]. Abdul M. et al a cancer prediction methodology utilizing nearest neighbor and support vector machine algorithms [11]. In [12], melanoma detection was accomplished using a support vector machine. Segmented images are used for the purpose of cancer classification. These ML strategies can only be used with the knowledge of dermatologists because they rely on crafted features. Finally, the need for human-engineered features is a hurdle for machine learning methods.

However, deep-learning techniques like convolutional neural networks (CNN) have gained popularity in the past decade due to their efficacy in automatic feature extraction and have been widely implemented by researchers. When it comes to classifying medical images, DL models have shown remarkable performance. Alizadeh et al. [13] developed a unique method for the automatic detection of skin cancer by employing a deep learning approach. They used an ensemble strategy to identify cancer by combining two CNN models with additional classifiers and extracting image texture features. Different evaluation metrics were then applied to the ISIC 2016, ISIC 2019, and PH2 datasets to evaluate the system. Zhang [16] developed a strategy for melanoma detection using a deep learning CNN model called EfficentNet-B6 that has been trained on the ISIC 2020 dataset. According to their claims, this model was used for the first time to identify skin cancer using transfer learning. Their system was evaluated using metrics from the area under the receiver operating characteristic curve. Yuan et al. segmented cutaneous lesions by utilizing a deep CNN [17]. In their study, Yu et al. employed highly intricate residual networks for the purpose of conducting automated identification of melanoma in dermoscopy images [18]. The authors Bi et al. utilized multi-stage fully convolutional neural networks to perform segmentation of dermoscopic images [19]. In their study, Dorj et al. utilized a convolutional neural network to classify skin cancer [20]. The authors Mahbood et al. conducted a skin lesion classification study using a fusion of deep neural networks [22].

All the above approaches rely on extensive amounts of dermoscopy images labeled with the right diagnosis as training data. To achieve adequate classification accuracy with less training data, several authors have developed methods. Currently, researchers frequently employ ensemble networks to enhance classification performance. Models are typically trained independently, and the final output is the result of applying majority voting and stacking techniques to combine the predictions of numerous models.

For skin lesion categorization, Aboulmira et al. [23] presented an ensemble network. Features are extracted using the separate models, and then the models are integrated to improve classification accuracy. The ISIC-2018 public data set was used to test the suggested ensemble of seven predictors, and the results show that it outperforms the state-of-the-art approaches. An enhanced capsule network (CapsNet) called FixCaps has been used for early skin cancer diagnosis in [24]. The proposed method not only enhanced its detection performance but also decreased the computational overhead by obtaining a bigger receptive field in comparison to the baseline CapsNet with a large kernel size of 31*31. Pacheco et al. [28] identified skin cancer using multiple deep learning models and information fusion method to combine metadata with pictures. They trained EfficientNet, DenseNet-121, MobileNet-v2, ResNet-50, and VGG-13 on two distinct datasets, specifically ISIC 2019 and PAD-UEFS-20, in along with the patient's medical information including age, gender, anatomical region cancer history, and skin prototype. Togacar et al. [30] introduced a unique model to identify skin cancer by employing the deep learning convolutional model MobileNetV2 combined with the spiking network to deliver the best results. Additionally, they reconstructed the dataset for training with an auto encoder model. The ISIC skin cancer dataset is used, which includes 1497 photos of malignant tumors and 1800 images of skin that is not cancerous.

Mahbod et al. [22] proposed a method for extracting deep features from multiple widely recognized and pre-trained models for the classification of melanoma. A multi-class SVM classifier was trained on features created by a pretrained AlexNet, ResNet-18, and VGG16 deep-feature generator. Finally, the results of the classifiers were combined to conduct classification. On the ISIC 2017 dataset, the suggested system achieved an area under the curve (AUC) of 97.55% for seborrheic keratosis (SK) classification and 83.83% for melanoma classification. Hosney et al. [31] describe a transfer learning-based method, which uses training data collected for a classification task other than skin cancer. Hosny's method is based on AlexNet model (a specific deep learning architecture proposed by Krizhevsky et al. [32]) that was initially trained to classify images on the ImageNet dataset and then modified to detect skin cancer through transfer learning.

The accuracy of skin cancer detection can be enhanced by combining the decisions of individual models. In this study, an ensemble of deep learners was built using VGG19, ResNet50 and MobileNetV2; the results indicate that the ensemble outperformed the individual learners. Based on previous research on classification approaches, it is observed that melanoma skin cancer detection model requires a hybrid deep learning method incorporating an ensemble technique.

III. PROPOSED METHODOLOGY

The proposed ensemble learning model is built in two phases using deep learning. We employed benign and malignant images from the International Skin Imaging Collaboration (ISIC) skin cancer images repository. In the initial phase, three pre-existing deep learning models, namely VGG19, ResNet50, and MobileNetV2, were developed. We used the stacking technique to combine the findings of each individual learner in the second stage. The diagram presented in the figure 3 illustrates the block diagram of the proposed approach for detecting melanoma skin cancer.

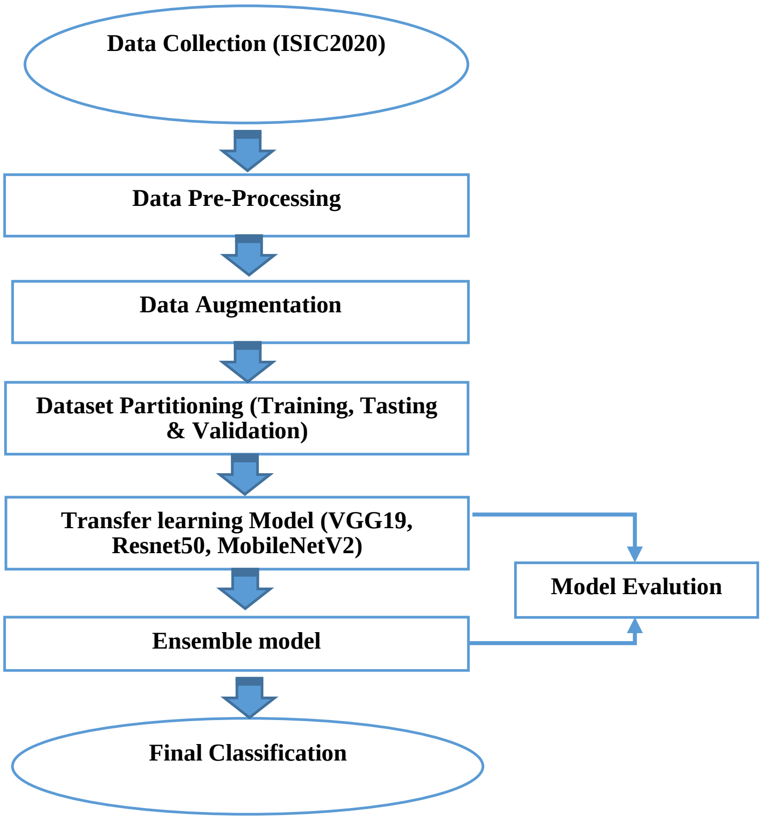

Figure 3: Proposed Methodology

A. *Dataset*

Data imbalance and data constraint are common problems in skin disease datasets, which might impede the categorization of skin cancer. It is significant to note that a substantial portion of data contained within various skin disease datasets pertains to non-malignant lesions. There are notable discrepancies in sample sizes among various categories of skin disorders in multiple datasets pertaining to dermatological conditions. We have used the ISIC 2020 dataset (https://www.kaggle.com/datasets/cdeotte/jpeg-melanoma-256x256?fbclid=IwAR3H1s3CG13zr7ba14rSHLHlNetIJKpT7CwNjduWz58L9JISae9aHwotbuo), which consisted of 32,126 skin lesion images and corresponding metadata encompassing the patient's identification, age and gender, lesion location, lesion classification, and malignancy status. The ISIC 2020 dataset is represented in Figure 4, which showcases a collection of images.

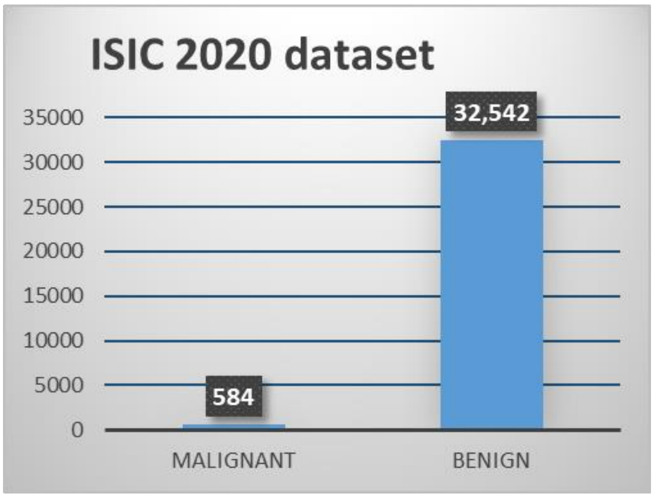

Figure 4: ISIC 2020 dataset

The dataset comprises 32,542 images portraying non-cancerous or benign skin lesions and 584 images portraying cancerous or malignant skin lesions.
The dataset demonstrates a notable imbalance, where a small proportion of 1.76% of the images portray malignant skin lesions, while the majority of 98.24% depict benign skin lesions.
To mitigate the issue of inadequate representation of malignant skin lesion images, the dataset underwent resampling. Resampling helps address class imbalance by generating balanced or modified datasets to improve model training. After resampling we got 2584 malignant (7.94%.) images, whereas the number of benign is 32542. The dataset also includes a CSV file with information about the patient, including the patient's id, gender, age, the location of any benign or malignant sites, and an indicative of malignancy for any imaging lesions.

B. Image Pre-Processing

Preprocessing is utilized on all input images of the ISIC-2020 dataset to enhance features and achieve greater consistency in classification outcomes. The convolutional neural network methodology necessitates a substantial quantity of iterative training. To mitigate the risk of overfitting, a voluminous image dataset was pre-processed. For using the pre-trained VGG19 and ResNet50 model, The input images are transformed from RGB to BGR, which subsequently zero-centers each color channel without scaling in relation to the ImageNet dataset. The original ISIC dataset comprises images that are obtainable in dimensions of 256 × 256. The dimensions of the dataset have been adjusted to 224 × 224. And for the MobileNetV2 the image size is like the other models, and we scale the input pixel within -1 to 1.

C. Data Augmentation

Data augmentation is a technique used to increase the size of a dataset by creating new data points through various transformations of the original data. This process can help improve the performance of machine learning models by providing more diverse and representative data for training.
The image data generator function of the Keras library in Python has been utilized to apply diverse data augmentation techniques. This has been done to address the issue of overfitting and enhance the variety of the dataset. The images were subjected to a rotation transformation in order to achieve a specific angle of rotation. Specifically, a rotation angle of 20◦ was employed for this purpose. The shear transformation method involves the fixation of one axis of an image, followed by the stretching of the other axis to a specific angle, commonly referred to as a shear angle. In this instance, a shear angle of 0.2 was utilized. The argument of zoom range was utilized to execute the random zoom transformation, whereby a value exceeding 1.0 denotes image magnification, while a value below 1.0 indicates image zoom-out. Consequently, the image was magnified using a zoom range of 0.95. Flip was utilized to horizontally and vertically invert the orientation of the picture. as presented in Table 1.

Table 1. Image augmentation techniques.

Transformations	Setting/Range
Rotation	20°
Horizontal flip	True
Vertical flip	True
Shear range	0.2
Zoom range	(0.95, 0.95)
Fill mode	Nearest

D. Training, Validation and Testing

We split the ISIC-2020 dataset into three sections: the training set, the test set, and the validation set. The training set was used to train the VGG19, ResNet50 and MobileNetV2 model, and the validation and test sets were then used to test how well it worked. So, we split the data set up as follows: 80% for training, 10% for testing, and 10% for validation. In the context of the ISIC-2020 dataset, the training, validation, and testing phases utilized 28100, 3513, and 3513images, respectively.

E. Deep Neural Network

To develop the proposed ensemble models, three deep network models, namely VGG19, ResNet50 and MobileNetV2 have been developed by fine-tuning the model parameters like batch size 64 and number of epochs 15. Short details of the models are described below.

1) Pre trained Transfer Learning Model:

Pre-trained transfer learning models refer to deep learning models that have undergone training on extensive datasets, usually comprising millions of labelled examples. The utilization of pre-trained transfer learning models offers several benefits such as decreased training time and computational resources, knowledge transfer and generalization, and enhanced performance on the intended task. Pre-trained transfer learning models are a significant repository of knowledge that can be fine-tuned on smaller target datasets. This approach enables the utilization of the learned representations and the attainment of competitive performance with limited data.

The size of skin cancer datasets is frequently restricted, posing a challenge for the development of deep learning models from the ground up. Pretrained models utilizing Convolutional Neural Networks (CNNs) acquire hierarchical representations of images, thereby facilitating the identification of significant visual patterns in skin cancer images.

2) *ImageNet*

The ImageNet dataset is a widely recognized and extensively utilized resource in the field of computer vision and deep learning research. In 2009, Fei-Fei Li and her colleagues at Stanford University established a standard for evaluating the effectiveness of image classification algorithms [32]. The dataset consists of a diverse collection of categorized images that cover a broad range of object classifications. Each image is accompanied by a corresponding label that identifies the object or scenery depicted in the image. ImageNet has undergone several expansions and updates over time, leading to an increase in its size and scope. The ILSVRC is a widely recognized benchmark that comprises approximately 1.4 million images representing 1,000 unique object categories.

3) *VGG19*

The VGG19 architecture is a convolutional neural network of significant depth that was formulated by the Visual Geometry Group at the University of Oxford. The architecture comprises a total of nineteen layers, comprising sixteen convolutional layers and three fully connected layers. VGG19 is advantageous due to its simplicity and uniformity, strong generalization capabilities, the ability to facilitate transfer learning, and its capacity for deep feature extraction. The VGG19 architecture has exhibited exceptional generalization abilities and is frequently employed as a foundational model for transfer learning. The process of deep feature extraction enables the extraction of intricate features from images across various scales, rendering it appropriate for intricate visual recognition tasks.

In our proposed model, we implemented our pre-trained VGG19 model using multiple techniques and configurations. We applied the following strategies and configurations so that we could achieve our performance goals: To prevent overfitting, early stopping with a patience value of 20 was employed. TensorBoard was utilized to visualize the training progress of the model. We used the ImageNet dataset to initialize the weights. We implemented Global Average Pooling2D for the pooling operation. To regularize the model's parameters, a kernel regularizer 'l2' penalty was used. ReLU and sigmoid activation functions were deployed. To prevent overfitting, we incorporated 256 neurons and a dropout rate of 0.4 into the model's first layer. We also used 128 neurons in the second layer with a dropout rate of 0.3 to improve the model's generalization skills.

4) *ResNet50*

The ResNet50 architecture is a convolutional neural network of significant depth, which was introduced by Kaiming He and colleagues in their scholarly publication "Deep Residual Learning for Image Recognition" in 2015. The structure comprises of 50 layers that are constructed using residual blocks. It follows a "bottleneck" design, wherein the 3x3 convolutional layers are flanked by 1x1 convolutional layers on either side. The benefits of this approach encompass a profound architecture that does not suffer from degradation, outstanding performance, and the capacity to address the issue of degradation that emerges as network depth increases. We implemented a ResNet50 model with early stopping (patience=20), TensorBoard monitoring, ImageNet weights, Global Average Pooling2D, 'l2' kernel regularization, ReLU and sigmoid activation functions. The initial layer consisted of 1024 neurons with a dropout rate of 0.3, followed by layers containing 512, 256, and 64 neurons, each with a dropout rate of 0.3. These configurations are intended to improve the performance and generalization capabilities of the model.

5) *MobileNetV2*

The architecture of MobileNetV2 is comprised of a sequence of depth wise separable convolutions, which effectively decrease the computational burden while preserving the ability to convey complex information. The methodology utilizes a fusion of 3x3 and 1x1 convolutional filters to capture both regional and global characteristics from the input images. Furthermore, skip connections are integrated to promote inter-layer information propagation, thereby enabling the model to effectively capture features at both low and high levels.

We used Early stopping with a patience of 20 and observed the training process with TensorBoard for implementing the pre-trained MobileNetV2 model. ImageNet was used to provide an initial value for each of the model's weights. We used a kernel regularizer with a 'l2' regularization, ReLU, and sigmoid activation functions in our model. Additionally, we added a Global Average Pooling2D layer. The first layer contained 256 neurons with a dropout rate of 0.4, whereas the second layer contained 128 neurons with a dropout rate of 0.3.

6) *Ensemble Learning:*

Ensemble learning is a concept that posits that the combined cognitive abilities of heterogeneous models can surpass the performance of a singular model, thereby enhancing precision and resilience. Ensemble learning has the potential to decrease errors, enhance generalization abilities, and furnish more dependable outcomes by amalgamating forecasts or judgements.

7) *Stacking:*

We ensembled our model by stacking. Stacking ensemble model combines predictions from multiple base models using a meta-model to improve overall predictive performance. We used the following three trees to ensemble our model.

The Random Forest algorithm employs bagging to select diverse subsets of the training data, ensuring sufficient dissimilarity among decision trees. In our study, we utilized 100 trees, each of which expanded down to a depth of 15 nodes. The splitting process persisted until a minimum of 30 records remained in a node.

In contrast, the Extra Trees algorithm utilizes the entire dataset to train decision trees. Similarly, we employed 100 trees, expanded down to a depth of 15 nodes, and performed splitting until at least 30 records remained in a node.

Both Random Forest and Extra Trees are decision tree-based methods. Decision trees are non-parametric and do not rely on assumptions regarding probability distributions. They exhibit robustness in handling high-dimensional data while maintaining good accuracy. Our experiment employed a maximum depth of 50 and a minimum number of samples required to split a node set at 50.

F. Different Metrics used to evaluate Performance

The following metrics were used to evaluate the performance of the suggested technique:

1) Accuracy

Accuracy is a performance metric for classifiers that is calculated as the proportion of correct predictions to the total number of predictions. It demonstrates how well the learning models can categorize the images.

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

where TP is true positive, FP is false positive, TN is true negative, and FN is false negative.

2) Precision

Precision is a performance measure that shows how accurately a classification model predicts the same result when a single sample is tested repeatedly. It evaluates the ability of the classifier to predict the positive class data samples. It is calculated as in Equation (2).

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

3) Recall / Sensitivity:

Recall is a classification measure that shows how many truly relevant results are returned. It reflects the ratio of all positive class data samples predicted as positive by the learner. It is calculated as in Equation (3).

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

4) Specify

To what extent a model is successful in identifying true negative cases is measured by its specificity. It is a metric for evaluating a model's ability to correctly identify instances that do not fit the target class. Specificity is computed as the proportion of true negatives to the sum of true negatives and false positives*:*

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

5) F-Score

F-score is calculated based on precision and recall. It can be considered as the weighted average of precision and recall. Its value ranges between [0, 1]. The best value of F-score is 1 and the worst is 0. It is computed as in Equation (5).

$$F\ Score = 2 * \frac{\mathrm{Precision*Recall}}{\mathrm{Precisio+Recall}} \qquad (5)$$

6) Confusion Matrix

The confusion matrix shows the correct and incorrect predictions made by a machine learning model and is used to evaluate the model's performance, as shown in Figure 5, a real positive can be found in the upper left corner, while a true negative can be found in the bottom right corner. Figure 8 depicts a false negative in the bottom left corner and a false positive in the top right.

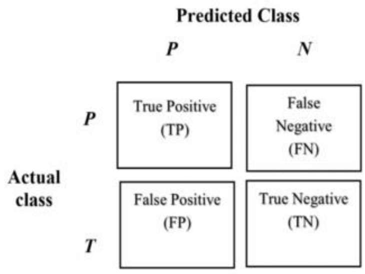

Figure5. Confusion Matrix

7) AUC Score and ROC Curve

The ROC curve represents a probability distribution, and its area under the curve (AUC) indicates the degree of separability. The ROC curve is a graphical representation that illustrates the relationship between sensitivity, which is the true positive rate, and specificity, which is the rate of false positives.

IV. RESULT & DISCUSSION

The performance analysis of the proposed ensemble model, along with individual deep learners is presented in Table 1.

Table 1. Comparison of Performance between ensemble model and the Individual learner.

Models	Accuracy	Precision	Recall	ROC-AUC	F1-Score
VGG19	91.54	91.54	1.2	49.93	1.9
ResNet50	99.57	97.64	96.51	98.16	99.81
MobileNetV2	91.09	44.91	94.18	92.51	91.98
Proposed Ensemble	99.40	94.73	97.29	98.43	95.99

According to Table 1, the accuracy values for VGG19, ResNet50, and MobileNetV2 are 91%, 99%, and 91%, respectively. The presented ensemble model has a prediction accuracy of 99%.

The recall values of VGG, ResNet50, and MobileNetV2 are 1%, 96%, and 94%, respectively, and 97% after ensembling as shown in Table 1.

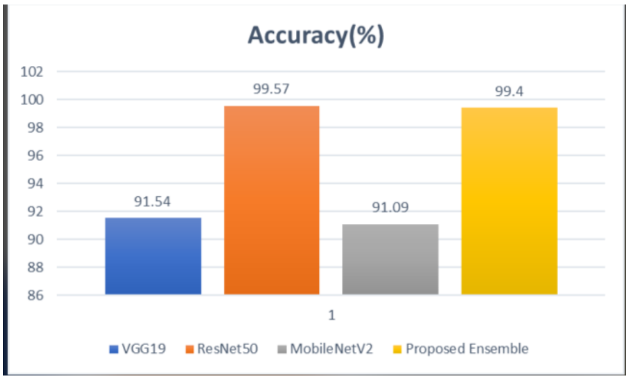

Figure6. The accuracy Comparison of the ensemble with other models.

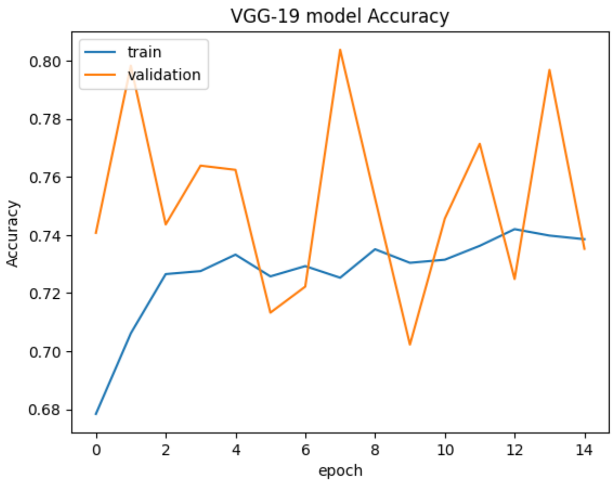

Figure7. Accuracy of VGG19

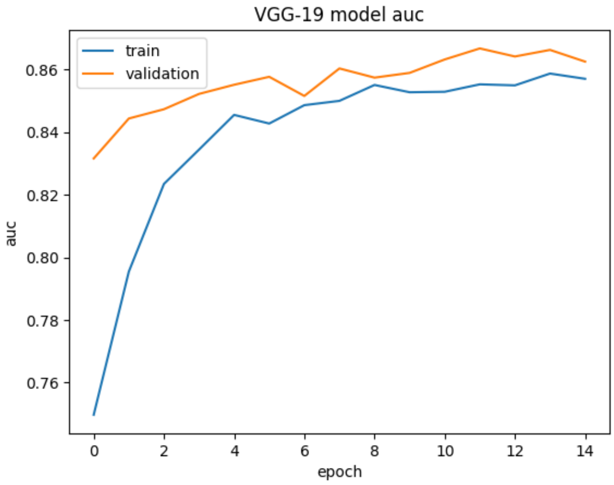

Figure8. AUC of VGG19

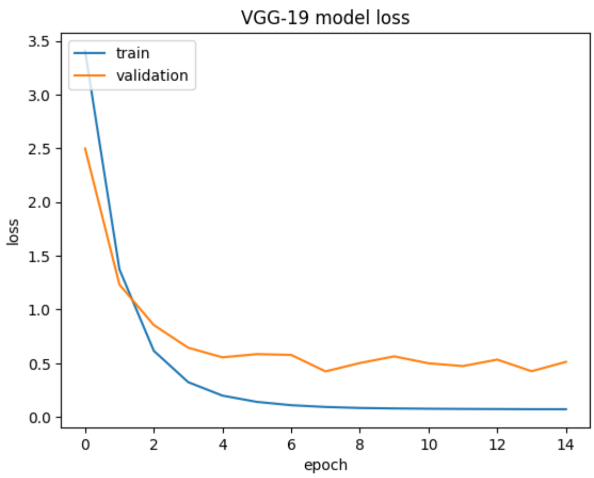

Figure9. Loss of VGG19

By incorporating these values, the proposed ensemble model classifies malignant images with greater precision than individual learners. The proposed ensemble system has a higher AUC value of 98% compared to the VGG, ResNet50, and MobileNetV2 (49%, 98%, and 92%, respectively), indicating that the proposed model has more accurately detected cancer through the combination of the decisions of individual learners and by taking advantage of their diversity. In addition, the table shows that the proposed ensemble model has a higher F1-Score at 95.99% compared to the individual learners, with values of 1.98%, 99 and 91% respectively.

The proposed model outperformed state-of-the-art methods in terms of various performance metrics such as accuracy, sensitivity, specificity, etc., as shown in Figure 9.
Individual learners' test and training accuracy are depicted in Figures 07, 11 and 15. Fig.8, 12 and 16 illustrate the AUC of the VGG19, ResNet50, and MobileNetV2. From the experiment our ensemble model accuracy is 99.4%. The figure indicates that the training accuracy of the ResNet50 model is 99. Figure 12 demonstrates that the MobileNetV2 model's training accuracy is 91%. Figures 9,13 and 17 depict the training and validation loss for each learner. According to the figures, the training loss for VGG19, ResNet50, and MobileNetV2 is 0.5, 0.4, and 0.2, respectively.

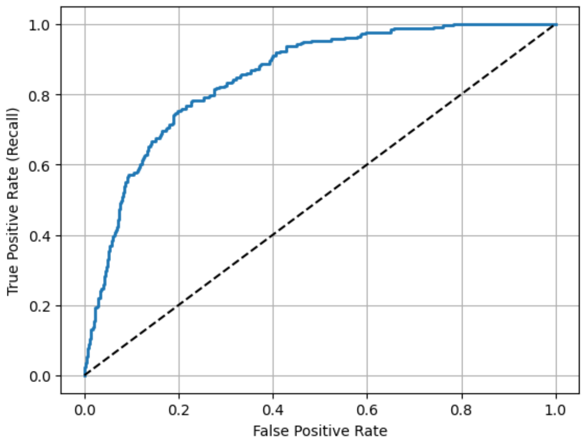

Figure10. ROC of VGG19

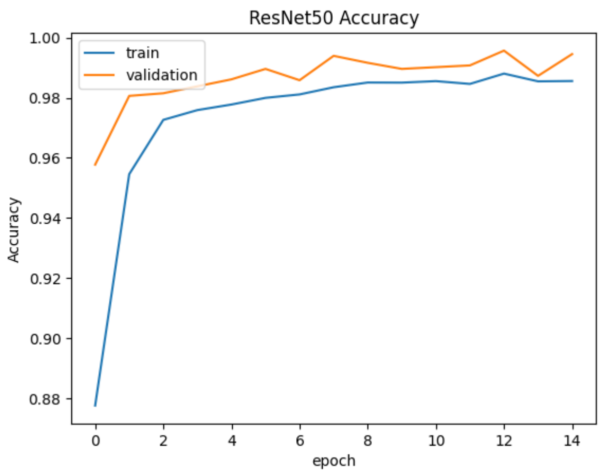

Figure11. Accuracy of ResNet50

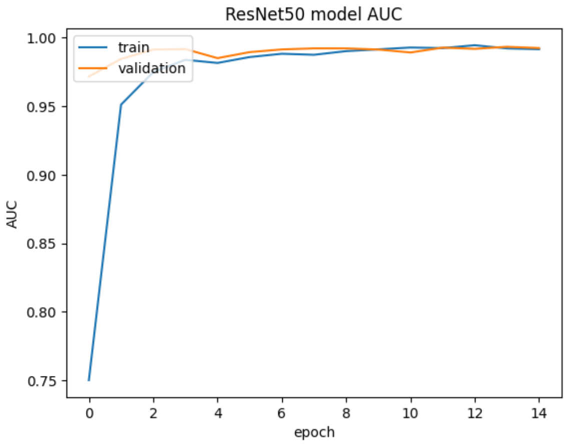

Figure12. AUC of ResNet50

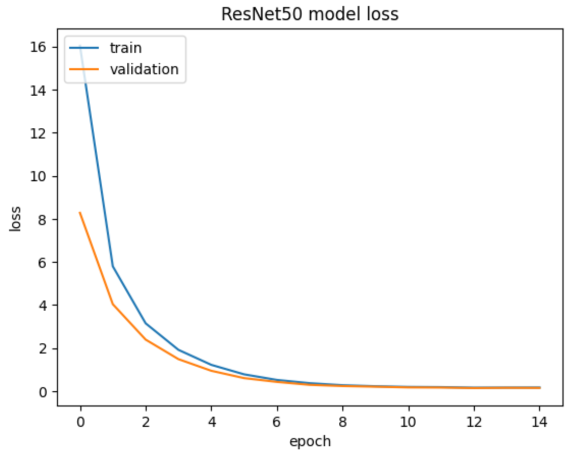

Figure13. Loss of ResNet50

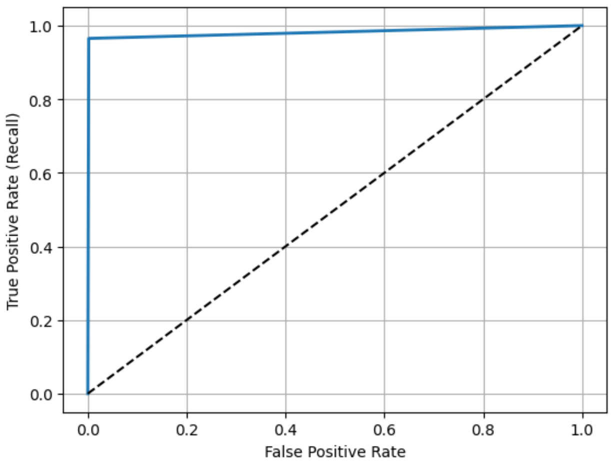

Figure14: ROC of ResNet50

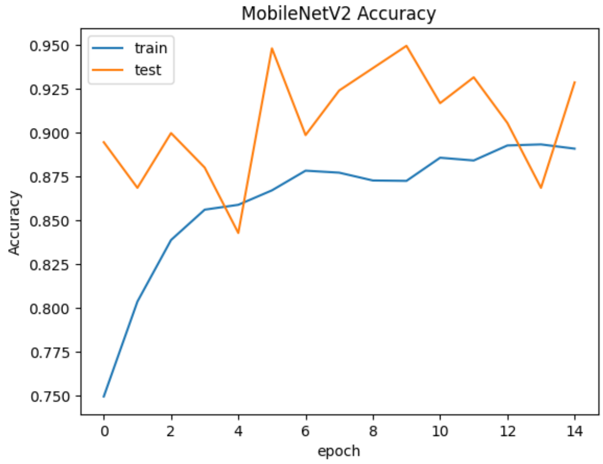

Figure15. Accuracy of MobileNetV2

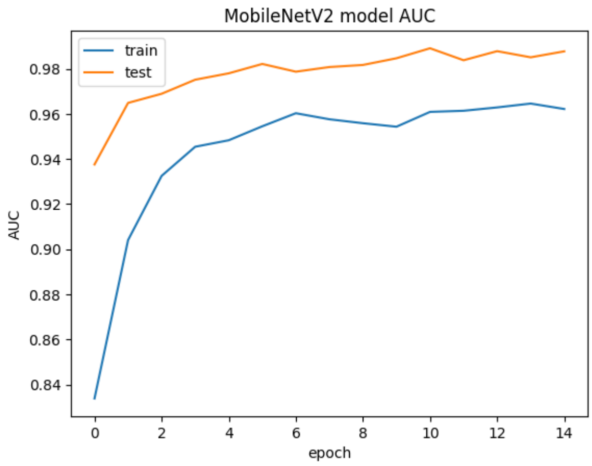

Figure16. AUC of MobileNetV2

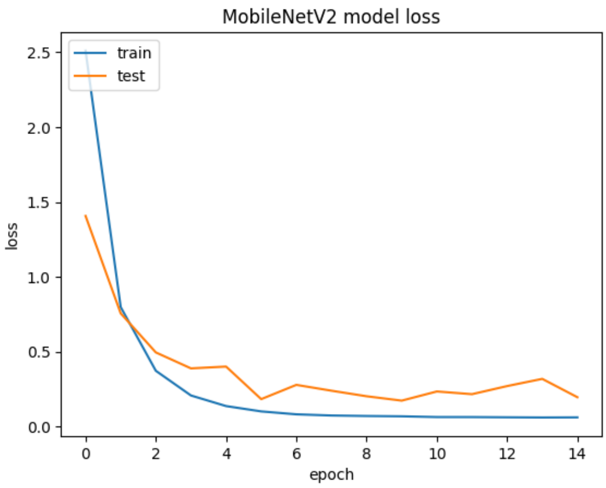

Figure17. Loss of MobileNetV2

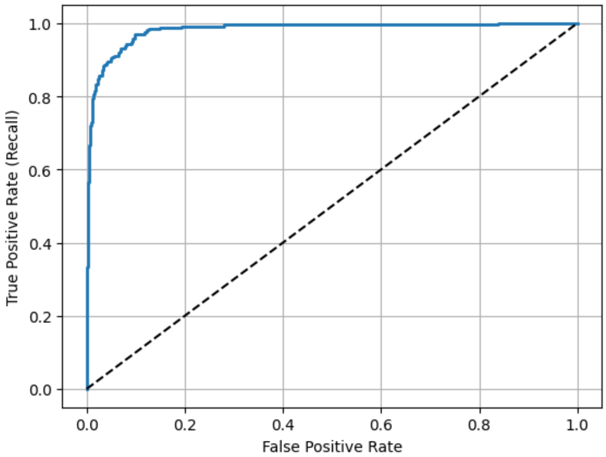

Figure18. ROC of MobileNetV2

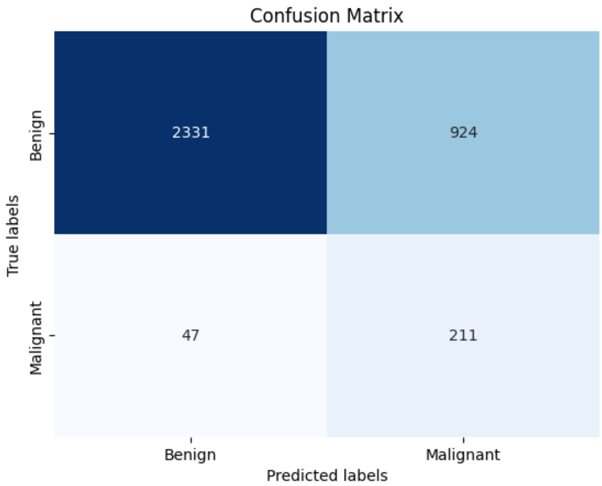

Figure19: Confusion metrics of VGG19

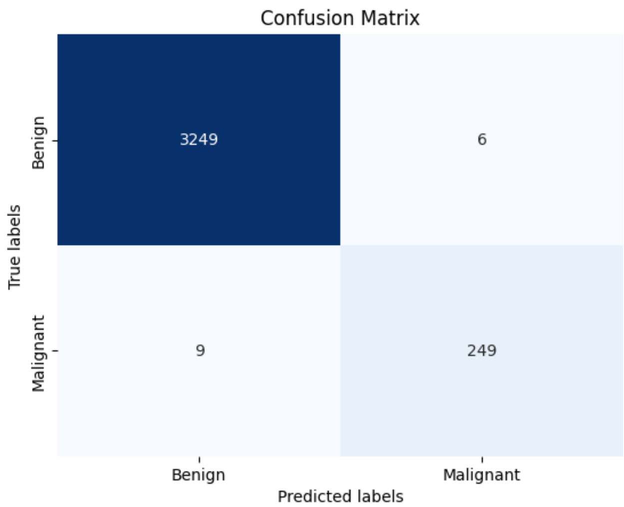

Figure20: Confusion metrics of ResNet50

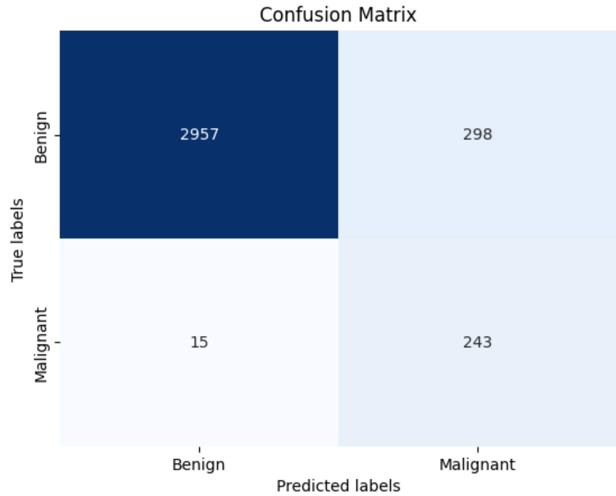

Figure21: Confusion metrics of MobileNetV2

A confusion matrix represents the truthfulness and falsity of an algorithm for machine learning. Figure 19-21 depicts the confusion matrices for VGG19, ResNet50 and MobileNetV2.

We have analyzed the performance of VGG19, ResNet50, MobileNetV2 and our proposed ensemble network. On the ISIC public dataset for skin cancer classification, the efficacy of the model is assessed. For the proposed model, the optimal hyperparameters, including 15 epochs, a learning rate of 0.0001 with a batch size of 64, and a stochastic gradient descent (SGD) optimizer, are determined using various batch sizes during the training evaluation. Figure 6 compares the accuracy of the proposed ensemble network to that of individual learners. Based on Fig. 6, it can be determined that the proposed model obtained the highest level of accuracy at 99.4%. The ROC curve of VGG19, ResNet50, and MobileNetV2 are depicted in Figures 10,14 and18.

V. CONCLUSION

Melanoma is the most dangerous form of skin cancer, but if diagnosed and treated in a timely manner, it may not be fatal. When it comes to making decisions on sensitive topics like cancer, it is possible to improve performance by merging the decisions of a variety of different individual learners. This paper describes the development of an ensemble model for detecting cutaneous cancer. It was developed by merging the decision-making capabilities of three different deep learning models: VGG19, ResNet50, and MobileNetV2. The proposed method has been implemented in the ISIC2020 challenge dataset of skin cancer photos to classify them as melanoma or not. Data augmentation techniques were employed to expand the dataset and enhance its precision. It is apparent from the findings that the proposed ensemble was successful in achieving an average accuracy of 99.4 percent. The suggested model achieves higher levels of performance than individual learners do in terms of sensitivity, accuracy, false-positive rate, specificity, and precision. In the future, we hope to investigate the efficacy of reinforcement learning-based methods in the identification of skin cancer.

REFERENCES