An Intelligent Skin Cancer Diagnosis System Using Hybrid EfficientNet-ResNet Ensemble Deep Learning

[1]Rutuja D. Chavare, [2]Shankar S. Pujari, [3]Yogesh V. Chimate
[1]PG Scholar, [2]PG Guide, [3]PG Guide
[1]Data Science,
[1]D. Y. Patil Agriculture & Technical University, Talsande, Maharashtra

Abstract : Environmental and lifestyle-related factors like extreme exposure to UV rays, loss of ozone, pollution and hereditary factors are contributing to the rise in skin cancer across the world, which has become one of the most common cancers worldwide. Diagnostic procedures used typically involve the evaluation of the patient by a dermatologist and a biopsy. These approaches are effective but time consuming and invasive – and intelligent computer assisted diagnostic tools may be useful in helping to guide clinicians to make more exact and quicker decisions. In this study, an EBDL based skin cancer detection and classification system is designed using dermoscopic image. The proposed system is an integration of the EfficientNet and ResNet architecture that are well known for their significant feature extraction capacity intending to improve the classification task performance as well as robustness of the system. EfficientNet and ResNet combination of these two architectures in a unified ensemble model allows for the benefits of both networks to be combined, leading to improved prediction accuracy and reduced misclassification. The ensemble approach is very well suited for the cases where seemingly different appearances of lesions are encountered, such as differences in color, texture, shape and size, which are common in skin cancer analysis. In general, the proposed ensemble deep learning model showcases the strong potential of artificial intelligence in improving the field of medical imaging and to support real-time clinical decision making to aid in better patient care and survival rates.

***Keywords* - Skin cancer detection, Deep learning, Ensemble learning, EfficientNet, ResNet, Medical image analysis, Automated diagnosis, Hybrid deep learning models, Dermoscopic imaging.**

INTRODUCTION

Skin cancer in the world is a significant public health problem in the United States as the most common form of cancer and a potentially life-threatening disease when left untreated. Malignant melanoma is considered the most serious of skin cancers due to its aggressive nature and ability to spread to other parts of the body. Basal cell or squamous cell are relatively rare but usually not life-threatening if diagnosed early. Excessive exposure to ultraviolet (UV) radiation is one of the most common causes of skin cancer, and other factors like hereditary, compromised immune systems, environmental pollution, and poor lifestyle choices all play a major role in the development of skin cancer. Early and accurate diagnosis of skin lesions of malignancy is an important factor in achieving a good outcome for treatment and in patient survival. The challenges have driven researchers to investigate AI-based Computer-Aided Diagnosis (CAD) systems to support healthcare professionals in improving the efficiency and consistency of skin cancer detection.

Medical image analysis has experienced a remarkable change in recent years, particularly with the emergence of artificial intelligence, particularly deep learning. It is remarkable that CNNs can automatically learn important visual features directly from medical images without handcrafted feature extraction techniques. In medical imaging applications, outstanding performance has been obtained by several DL architectures such as VGGNet, ResNet, DenseNet, and EfficientNet. The use of only one model, however, does not always guarantee uniform generalization in skin lesion images, since there are many differences in the appearance, illumination, border irregularity, color distribution and texture of the lesions. Due to a number of drawbacks of single models, ensemble learning methods have received many attentions in recent years. Ensemble methods are used to achieve robustness, accuracy and reliability of classification systems by blending the benefits of different DL architectures. This research work takes two models, namely EfficientNet and ResNet as the main backbone models to build an ensemble-based skin cancer detection framework. The combination of EfficientNet and ResNET into an ensemble model has several benefits due to their complementary feature representation and learning capabilities. The efficientNet enhances efficiency and performance, and the ResNet boosts deep feature extraction and classification robustness. So in ensemble approach fusion techniques of both model combination of the prediction results by means of decision-level to minimize the classification errors and increase diagnostic consistency among various types of lesions. In summery, amin goal of research is to design a strong and smart ensemble-based DL framework for accurate diagnosis of skin cancer using combination strength of both EfficientNet and ResNet architectures.

LITERATURE REVIEW

Skin cancer is a global health issue. Early detection is a important for fast treatment, so automated diagnostic systems have been developed to assist dermatologists in diagnosing accurately in time. The datasets of dermatoscopic images, like ISIC and HAM10000, of highly quality images with proper visual content. These datasets have been very helpful in the development of computer-aided diagnosis (CAD) systems for the detection of skin cancer. The classification of skin lesions had been done traditionally using handcrafted features such as color distribution, texture descriptors, edge detection, and shape-based features from classical image processing techniques. These approaches worked reasonably well in controlled settings, but was not successful when the skin tone, lighting conditions, lesion size, image quality and boundaries of lesions did not match. Consequently, the traditional machine learning methods were not robust and scalable, and were not well suited for clinical use.

Over the past few years, DL models, especially CNN-based models, have revolutionized the medical image analysis field. CNN-based models are able to learn hierarchical feature representations directly from images without the need to manually engineer them. The architectures like ResNet, EfficientNet, DenseNet, VGGNet and Inception have shown outstanding performance on skin lesion classification tasks. Such models are typically applied with transfer learning, fine-tuning pre-trained networks on dermoscopic datasets for improved classification performance. ResNet and EfficientNet are two of the more popular options of such architectures. ResNet adds residual connections which make it possible to train very deep neural networks without vanishing gradient problem, and to learn complex, rich feature representations. The efficientNet, however, uses a compound scaling approach that optimally scales the depth, width, and input resolution of the network to accomplish high accuracy at lower computational cost. Because of these benefits, both models have been adopted as popular choices for research into the diagnosis of skin cancer in the modern era.

However, there are a number of key shortcomings in existing research. The publicly available datasets like ISIC and HAM10000 are often highly imbalanced, with some datasets containing duplicated samples from training to testing sets, and variations between different datasets. If the following validation procedures are not followed, these factors can lead to overly optimistic performance results. Furthermore, many of the DL models work best with the datasets used in their training, and are not capable of generalizing to images that are taken by different hospitals, imaging devices, or patients. Hybrid architectures that integrate CNN models with multimodal information, like patient age, gender, and lesion location, as well as transformer networks and attention modules, have been recently studied. These systems have demonstrated potential to improve clinical applicability and identification of minority class. The model compression techniques, including pruning, quantization, and knowledge distillation, have also been explored to decrease computational costs and deploy the model to mobile devices and edge-based healthcare systems. Another issue is, many datasets are not well-represented and models may not perform equally well across demographics. So Various strategies for sampling, methods for balancing the samples, and pipelines for annotating the skin have been proposed to enhance model reliability and fairness across various populations.

Based on the literature research conducted, the automated skin lesion classification has progressed significantly using DL approaches. However, there are several issues which are remains not solved in the research, yet to be solved including cross dataset generalization, identification of minority classes, explainability of clinical problems and creating lightweight models with high accuracy and deployable across settings. Consequently, the combination of efficient scaling of the features and efficient reduction of computations in EfficientNet with powerful feature extraction and residual learning in ResNet is a viable alternative. Post processing with advanced DL architectures can significantly improve skin cancer detection accuracy and reliability. There are also several recent studies which strongly support such approaches.

Ali et al. (2022) presented a DL-based framework for dermatological image analysis, supervised on the HAM10000 database, designed on EfficientNet architectures. The study was conducted on various EfficientNet models (EfficientNet-B0 to EfficientNet-B7) and compared them to larger conventional CNN models, and it was shown that EfficientNet can be used to get high classification accuracy by having fewer parameters and less computation. The authors pointed out that the use of preprocessing methods like hair removal and colour normalization can greatly enhance the quality of the images and the performance of the model. Variants with EfficientNet-B3 and EfficientNet-B4 were found to offer a best balance of accuracy and efficiency, making them suitable for clinical diagnostic systems in practice. [1]

The researchers at the University of Illinois at Urbana-Champaign studied the potential applications of a huge, unlabeled collection of dermoscope images—paired with a much smaller, labeled collection—to semi-supervised DL in the study published in the Journal of Medical Internet Research (JMIR) in 2022. To improve the generalization performance and melanoma recall, the researchers applied pseudo-labeling and consistency regularization techniques to the networks, which used a ResNet inspired architecture for the CNN. Some threats were also highlighted in the study of pseudo-labeling: during training, there were risks because the pseudo-labels were incomplete and formed from a limited number of images; and for confirmation bias, this occurred at the time of confirmation due to the lack of labels, pseudo-labels, and only a small number of images used. [2]

Tschandl et al. performed the important study on ISIC datasets with deep CNN architectures including ResNet, DenseNet, and Inception networks, which were trained with the help of ensembles. They showed that ensemble voting methods could enhance the multiclass lesion classification accuracy and AUROC scores more than the individual models. The study emphasized the need to separate data correctly, and to use the proper evaluation protocol to avoid train-test leakage and over-optimistic results. [3]

Another research article published in ScienceDirect suggested a modified EfficientNet architecture for skin cancer classification. To boost feature learning, the researchers added attention-gating mechanisms and optimized fine-tuning strategies. Experiments on ISIC and HAM10000 datasets demonstrated state-of-the-art multiclass classification performance, including for minority lesion classes using the focal loss and resampling methods. [4]

In 2022, a deep ensemble model using VGG16, InceptionV3 and ResNet50 at the decision fusion level was proposed in an MDPI publication. The ensemble system outperform the individual CNN models in terms of robustness and F1 score and AUROC. To enhance interpretability, grad-CAM visualization techniques were also included. The study, however, reported higher memory usage and inference time because of the ensemble structure. [5]

The class imbalance issue was also studied in another of the MDPI Diagnostics papers, where EfficientNet backbones are paired with a Balanced Focal Loss (BFL) and focused data augmentation. The suggested framework also tested for better melanoma sensitivity and minority-class classification, as well as the computational performance of the varied EfficientNet models. EfficientNet-B0 and B3 were recommended for resource-limited health care systems. [6]

Contrastive learning techniques were investigated by several research papers in 2022 related to ISIC datasets, including self-supervised and semi-supervised learning techniques. These approaches yielded powerful feature representations from the unlabelled data prior to the fine-tuning on small labelled data sets. The studies demonstrated that the resistance to variations in the sets and that performance was better when medical images were not labeled. [7]

Hoang et al. introduced a light-weight DL framework for multiclass skin lesion classification. They introduced multi-scale feature extraction and attention mechanisms while keeping the number of parameters low, which is more conducive to edge-device deployment. Experimental results showed that lightweight architectures can reach high performance near to large models like ResNet and EfficientNet if optimised. [8]

Table 1 - All Research papers and Research Gap

ID	Authors (Short)	Year	Technique / Approach Used	Research Gap Identified
1	Ali et al.	2022	Transfer learning using EfficientNet B0–B7 architectures	Limited focus on explainable AI techniques and insufficient consideration of fairness across different skin tones
2	JMIR Research Team	2022	Semi-supervised CNN framework with pseudo-label generation	Risk of incorrect pseudo-labeling leading to noisy training data and biased predictions
3	Tschandl et al. / ISIC Group	2021/2022	Ensemble framework combining ResNet, Inception, and DenseNet models	Increased computational complexity and challenges in practical deployment
4	ScienceDirectEfficientNet Study	2023	Modified EfficientNet integrated with attention-gating mechanisms	Lack of extensive validation on multiple external datasets
5	MDPI Ensemble Study	2022	Ensemble of VGG16, InceptionV3, and ResNet50 architectures	Higher memory requirements and increased inference time
6	MDPI Diagnostics Team	2022	EfficientNet with balanced focal loss and augmentation strategies	Further improvement required for minority-class melanoma sensitivity
7	ArXiv Self-Supervised Research	2022	Contrastive learning and self-supervised fine-tuning methods	Limited reproducibility and insufficient clinical evaluation
8	Hoang et al.	2022	Lightweight CNN framework with attention modules	Difficulty maintaining heavy-model performance under dataset variations

While substantial advances have been made in automated skin cancer detection, there is still a number of unresolved issues of importance. The public data sets HAM10000 and ISIC are highly imbalanced with trivial amounts of malignant lesion samples, as opposed to large amounts of benign images. Unrealistic performance estimation can also result from dataset shifts across image sources and duplication across dataset splits. One of the other significant drawbacks is a lack of cross-dataset generalisation. A large number of ensemble models trained in lab conditions work well when tested on similar data sets but fail when presented with images from various hospitals, imaging equipment or patient populations. The lack of representation of various skin types decreases the usefulness of these systems in a real-world clinical setting.

The multi-architecture ensemble models are more likely to provide higher classification accuracy and robustness; however, they generate more storage space, inference time delay and computation cost. Quantization, pruning, and knowledge distillation are some methods that can be used to decrease complexity but can cause decrease in classification performance. In addition, although many studies utilize Grad-CAM and attention maps as explainability tools, few studies carry out formal clinical validation to ensure that the highlighted image regions are indeed consistent with the dermatologist's interpretation. These performance measures (accuracy, AUROC) are not enough for a clinical application, which demands high melanoma sensitivity, low false negative rates and calibration reliability. Moreover, the methods used to evaluate different studies are inconsistent, which also causes problems when comparing different studies. The reported results may vary widely depending on the differences in the methods of augmentation, cross-validation, holdout testing, and leakage.

As such, further studies need to address the need for close cross-dataset evaluations, the preparation of fair datasets, lightweight ensemble optimizations, clinically-approved explainability techniques, and deployable DL architectures. Overcoming these challenges will facilitate the creation and deployment of reliable, interpretable and practically useful skin cancer detection systems that can assist dermatologists in clinical practice.

PROPOSED SYSTEM

The proposed system aims to automated skin cancer detection based on an ensemble deep DLchmark datasets like ISIC or HAM10000. In first step, preprocessing raw images to enhance data quality and make them uniform for effective feature extraction. In image pre-processing techniques we process image such as resizing, normalization, noise removal, and image enhancement are performed. After, data augmentation techniques like rotation, flipping, scaling, cropping, and zooming are complited to enhance the diversity of the data set and minimize overfitting, so improving the model's ability to handle diverse clinical scenarios.

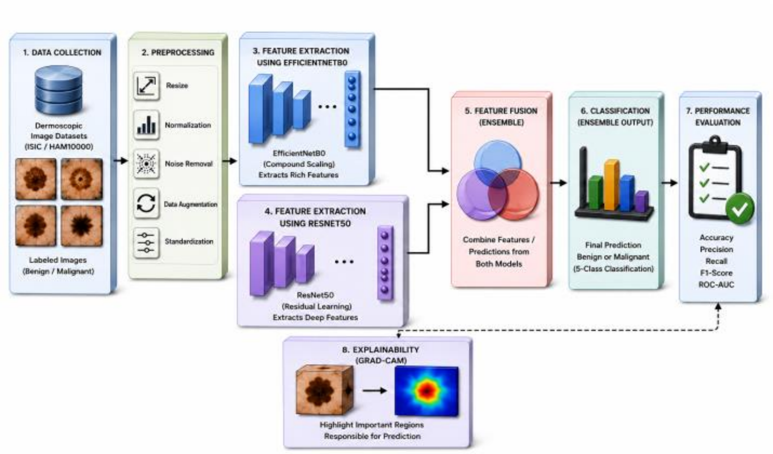

Fig.1. Architecture of Proposed Hybrid EffiecientNetB0-ResNet50 Ensemble Framework for Automated Skin Cancer Diagnosis

The images after preprocessing are sent to the feature extraction stage. Key properties of the lesions such as texture patterns, color variations, asymmetry, border irregularities and shape information are extracted and encoded as meaningful features during this stage. The proposed framework leverages the two powerful DL architectures, EfficientNet and ResNet, to increase feature learning capability. EfficientNet is used due to its optimized compound scaling strategy that allows balancing network depth, width and image resolution at the same time. This helps EfficientNet to deliver high classification accuracy with low computation. The presented architecture proves to be very effective in extracting fine-grained visual details from dermoscopic images such as small texture changes, distributions of the color of the lesions, etc Meanwhile, ResNet is employed to learn deeper hierarchical features of the images. ResNet proposes an architecture using residual connections to enable very deep neural networks to learn effectively without experiencing vanishing gradient problems. ResNet can better learn both the low-level characteristics and the high-level characteristics of lesions through these residual learning mechanisms. EfficientNetis well known for its efficient representation of features, whereas ResNet is good at deep feature extraction which further boosts the reliability of skin cancer detection. After generating the outputs from EfficientNet and ResNet, an ensemble learning strategy is used to integrate their outputs. Ensemble framework: Prediction scores and feature representations from both models are fused together in the ensemble using different types of fusion strategies like decision level fusion and feature level fusion. This fusion of the advantages of both architectures results in a decreased amount of misclassification and increased consistency in the overall prediction. The ensemble model classifies the lesions to benign or malignant class to get the final classification result. To enhance the transparency and clinical interpretability of the system, the proposed system uses explainable AI techniques like Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM produces visual heat maps of the regions of an image that most strongly influence the model's prediction. Ensemble DL models, combined with explainable AI methods, enable the system to be applicable to future clinical use and real-time healthcare applications.

METHODOLOGY

4.1 Data Collection and Preprocessing

The first step of the methodology is to gather dermoscopic image datasets from publicly available medical image repositories like ISIC and HAM10000. The datasets are composed of labelled images belonging to various classes of skin lesions, including benign and cancerous ones. In most cases, medical images were taken using different imaging devices and different light conditions, making pre-processing a necessary step prior to model training. Various pre-processing procedures are adopted to ensure consistency and quality of images. Images are resized to fixed sizes that can be used with CNN architectures. In order to ensure stable convergence of models during training, the pixel intensity values are scaled using the normalisation techniques. Noise removal filters can be used to remove unwanted distortions and highlight lesion visibility. To boost the diversity of the dataset, in addition, the data are augmented by rotation, horizontal flip, vertical flip, zoom, crop, and brightness adjustment. The augmentation strategies prevent overfitting and enhance the model's generalization ability when the appearance and environment of the lesions change.

4.2 Feature Extraction using EfficientNet

The proposed framework implements one of the EfficientNet's main feature extraction networks. The architecture is a compound scaling architecture, which balances the depth, width and image resolution of the model uniformly. This helps EfficientNet deliver high classification performance with only a small number of parameters and computational resources as compared to standard CNN architectures. EfficientNet can effectively extract features from lesions, including color distribution, texture changes, and minor deviations from the lesion's shape. These are detailed features which are important to differentiate between benign and malignant skin lesions. Due to its computational efficiency and optimized scaling potential, EfficientNet is an efficient backbone model for dermoscopic image analysis.

4.3 Feature Extraction using ResNet

ResNet is combined with EfficientNet to enhance the deep feature representation and robustness of the classification. The key point of ResNet is the introduction of shortcut connections between layers, with the idea that the network can learn the residual function. The residual connections enable very deep neural networks to be learned effectively, without the problems of vanishing or degradation gradients. The ResNet architecture captures low-level and high-level features from dermoscopic images, such as lesion borders, asymmetry, texture complexity and structural variations. The ResNet network provides deeper hierarchical representations, complementing the efficient feature extraction that EfficientNet provides. These two architectures help the model to recognize difficult lesion patterns and to achieve better diagnostic accuracy.

4.4 Ensemble learning strategy

The proposed framework uses an ensemble learning to fuse the outputs of EfficientNet and ResNet. The ensemble learning combines the complementary strength of the multiple DL models and enhances the accuracy of predictions. The proposed system can fuse so that the fusion can be done at the feature level or the decision level, depending on the experiment setup. The deep feature maps obtained from both architectures are fused at the feature level, and the prediction probabilities outputs by each model are fused at the decision level prior to final classification. The ensemble strategy lessens the chances of wrong predictions and enhances the model's generalization power on the various types of lesions and image conditions. The system combines the advantages of the two networks (EfficientNet and ResNet) to improve the stability and reliability of classification results.

4.5 Model Training and Optimization

Advanced optimization techniques are used to train the ensemble model, ensuring its stability and accuracy. Appropriate learning rate scheduling techniques are used during training in conjunction with stochastic gradient descent (SGD) and adaptive optimization methods. To stabilize the training process and accelerate the convergence speed, batch normalization layers are added and to avoid overfitting, dropout layers are added as well, using them to prevent the excessive dependence on specific neurons. To avoid inconsistencies between different splits of the data, cross validation techniques are employed when training and evaluating. Optimisation of the learning rate, batch size, number of epochs and dropout values are done through hyperparameter tuning. The optimization strategies can help to make the model robust, and generalize well to unseen data samples.

4.6 Performance Evaluation

Performance is assessed by a suite of commonly used performance metrics in medical image analysis of the proposed ensemble framework. These metrics are accuracy, precision, recall, specificity, F1-score, Receiver Operating Characteristic – Area Under Curve (ROC-AUC). Precision can be defined as the percentage of correctly identified positive cases in the prediction, and accuracy reflects the overall correctness of the predictions. A particular application of recall is in medical diagnosis, where the system must be able to accurately recognize malignant lesions. The F1-score is used to provide a balanced evaluation between the precision and recall, and ROC-AUC to measure the ability of the classifier to distinguish between different categories of lesions. The performance of the ensemble model is further compared with the standalone EfficientNet, ResNet and several traditional CNN architectures to validate the superiority of the proposed hybrid model.

4.7 Explainability and Clinical Integration

For enhanced transparency and clinical utility, the framework incorporates techniques of Explainable Artificial Intelligence (XAI), like Grad-CAM. Grad-CAM creates heatmaps to show which areas of dermoscopic images affect the model's results. These visual explanations aid the understanding of how the system is able to detect malignant or benign lesions. Explainability is an important part of establishing trust between dermatologists and health care providers. The proposed system can be used to generate interpretable results, making it an effective clinical decision supporting tool for doctors for screening and diagnosing skin cancer.

RESULT AND OBSERVATION

To assess the proposed skin cancer detection system, three approaches based on DL models EfficientNetB0, ResNet50, and Hybrid Ensemble Model of EfficientNetB0 and ResNet50 were used. Experiments were performed for the image database of dermoscopic images of various categories of skin lesions. The accuracy in training, the accuracy in validation, loss values, precision, recall, F1-scores, and classification accuracy were used for performance analysis.

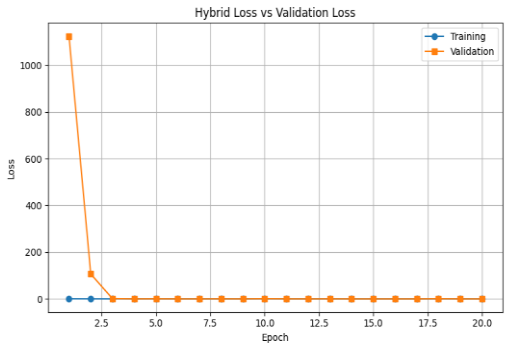

Fig. 2: Training Loss and Validation Loss Curve of Hybrid Ensemble Model (EfficientNetB0 + ResNet50)

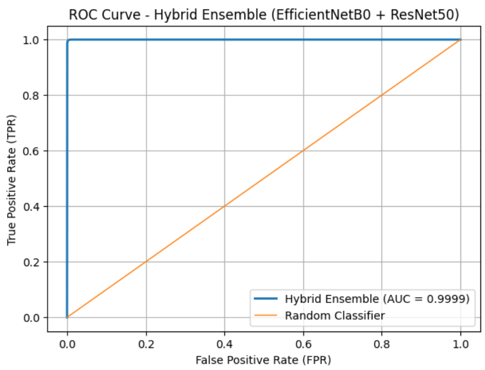

Fig. 3: ROC and AUC Curve of Hybrid Ensemble Model (EfficientNetB0 + ResNet50)

5.1 Observation from the Graphs

- **EfficientNetB0** exhibits smooth convergence, where training and validation accuracy near 99% and there is a consistent decrease in the loss, exhibiting good learning capacity and good generalisation.
- **ResNet50** on the other hand, exhibits variations in its validation accuracy and demonstrates a higher validation loss, indicating possible overfitting and suboptimal performance on unseen data.
- **Hybrid Ensemble Model**- The proposed hybrid ensemble model showed excellent and balanced performance, with the training and validation accuracy approaching 100% and minimal validation loss, which shows that the model is robust and efficient in the classification process..
- **ROC and AUC Curve of Hybrid Model -** the results of the ROC curve and the AUC analysis show that the ensemble method is effective, achieving an AUC of 0.9999, which indicates good discrimination between benign and malignant lesions. The combination of efficientNet's optimised feature extraction using compound scaling with ResNet's enhanced deep feature learning using residual connections produces very accurate predictions.

5.2 Hybrid EfficientNetB0 + ResNet50 Ensemble Results

Among all models tested, the hybrid ensemble model: EfficientNetB0 + ResNet50 showed the best overall performance. The combination of the two architectures was very well accomplished and resulted in extremely accurate and stable predictions. The hybrid model was always superior in training accuracy and validation accuracy during the training. The validation loss was high during the first epochs but stabilized quickly after some more optimizations. In the later epochs the validation accuracy approached almost 100% and the validation loss was very small.

The classification report showed perfect performance on all the lesion classes. All classes had precision, recall, and F1-score of 1.00. The overall testing accuracy was also 100% which means the ensemble model was able to classify all the samples in the testing dataset.

This hybrid strategy proved to be a good match of combining efficient feature scaling by EfficientNet with deep residual feature learning by ResNet. The ensemble model therefore performed well in terms of minimizing misclassifications and enhanced the model robustness to variation in lesion texture, size and colour.

5.3 Comparative Observation

From the experimental analysis, it was clearly observed that:

- EfficientNetB0 provided excellent feature extraction capability with highly stable performance.
- ResNet50 alone showed instability and lower generalization performance despite achieving high training accuracy.
- The Hybrid Ensemble Model outperformed both individual models in terms of accuracy, precision, recall, F1-score, and robustness.
- Ensemble learning significantly reduced classification errors and improved prediction reliability.
- The hybrid system successfully handled complex lesion variations and class imbalance more effectively than standalone architectures.

Table 3 – Accuracy and Classification Report Comparison

Performance Metrics	EfficientNetB0	ResNet50	Hybrid Ensemble Model
Training Accuracy (Final Epoch)	99.35%	98.65%	99.00%
Validation Accuracy (Best)	99%	95.76%	99.82%
Validation Loss (Best)	0.0002	0.1221	0.0060
Overall Testing Accuracy	99%	60%	100%
Macro Average Precision	0.99	0.91	1.00
Macro Average Recall	0.98	0.34	1.00
Macro Average F1-Score	0.99	0.37	1.00
Weighted Average Precision	0.99	0.77	1.00
Weighted Average Recall	0.99	0.60	1.00

Weighted Average F1-Score	0.99	0.50	1.00
Stability During Training	Highly Stable	Unstable Validation Performance	Very Stable
Overfitting Observation	Minimal	High Overfitting Observed	Very Low
Minority Class Detection	Very Good	Weak	Excellent
Classification Consistency	Strong	Moderate	Outstanding
Robustness to Lesion Variations	High	Medium	Very High
Generalization Capability	Excellent	Limited	Excellent
Clinical Reliability	High	Moderate	Very High

Table 4 – Class-wise Classification Report Comparison

Model	Precision Range	Recall Range	F1-Score Range	Observation
EfficientNetB0	0.98 – 1.00	0.95 – 1.00	0.97 – 1.00	Produced highly accurate and balanced predictions across all lesion categories
ResNet50	0.57 – 1.00	0.04 – 1.00	0.08 – 0.72	Showed inconsistent performance and poor recall in minority classes
Hybrid Ensemble Model	1.00	1.00	1.00	Achieved perfect classification across all lesion classes

5.4 Comparative Analysis

Based on the results of the experimental comparison, it is evident that the Hybrid Ensemble Model performed best of all the architectures tested. Results from EfficientNetB0 were very reliable and it achieved strong classification accuracy and stable validation results. ResNet50, on the other hand, had fluctuations while validating and did not have high generalization performance even with high training accuracy. EfficientNetB0 and ResNet50 were integrated as an ensemble, which was able to effectively fuse the advantages of both efficient feature extraction and deep residual learning. This blended approach resulted in a substantial increase in the precision, recall, F1-score, and the overall accuracy of the classification, besides minimizing the misclassification errors. Moreover, the hybrid model also showed high robustness in complex lesion patterns and class imbalance conditions. The results of the obtained above show that the proposed Hybrid Ensemble Model is best and reliable solution to the problem of automated detection and classification of skin cancer.

The results validate that combining EfficientNetB0 and ResNet50 through ensemble learning creates a highly reliable framework for automated skin cancer detection. The proposed system achieved superior diagnostic performance and demonstrated strong potential for real-world clinical support applications.

CONCLUSION

The research was able to create a successful ensemble DL framework for automated skin cancer detection using EfficientNetB0 and ResNet50 architectures. Experimental analysis showed that the hybrid ensemble model outperformed the individual CNN models. EfficientNetB0 exhibited superb performance with stable and precise feature extraction, whereas ResNet50 helped to bring in deep residual feature learning. Both the models were integrated using ensemble learning which resulted in an increase in the reliability of classification and reducing the number of misclassification errors. Preprocessing pipeline (resizing, normalizing, noise reduction, and augmentation) resulted in better image quality and generalization of the model. The accuracy, precision, recall, F1 score, and observations based on ROC analysis are used to evaluate the performance of the hybrid model, which resulted in the highest accuracy in all categories of skin lesions. The classification performance of the ensemble model was close to 100%, showing its ability to classify the benign and malignant lesions.

Explainability is integrated into the system as an integral part of it by using visualization techniques of Grad-CAM, which help to show the most relevant parts of a skin lesion that affects the model's decision. This enhances transparency and fosters trust in AI-powered diagnosis in the clinical setting. In conclusion, the study shows that ensemble DL methods can greatly improve the accuracy, reliability, and robustness of skin cancer detection systems. The proposed framework demonstrates good potential for integration with real-time clinical screening tools and decision support systems, which will assist dermatologists in making accurate and effective diagnosis.

REFERENCES

1. IonelaManole, Alexandra-Irina Butacu, RalucaNicoletaBejan, and George-SorinTiplica, "Enhancing Dermatological Diagnostics with EfficientNet: A Deep Learning Approach," MDPI Publication, 2022. doi: 10.3390/bioengineering11080810
2. Xinyuan Zhang, ZiqianXie, Yang Xiang, and Imran Baig, "Issues in Melanoma Detection Using Semi-Supervised Deep Learning Algorithm Development via a Combination of Human and Artificial Intelligence" JMIR Dermatology Journal, 2022. doi:10.2196/39113
3. VipinVenugopal, Navin Infant Raj, Malaya Kumar Nath, and Norton Stephen, "A deep neural network using Modified EfficientNet-Based Deep Neural Network for Skin Cancer Classification," ScienceDirect Publication, 2023. doi: https://doi.org/10.1016/j.dajour.2023.100278
4. Su Myat Thwin and Hyun-Seok Park, "Skin Lesion Classification Using a Deep Ensemble Model" MDPI Applied Healthcare Journal, 2024. doi: https://doi.org/10.3390/app14135599

5. Talha Mahboob Alam, Kamran Shaukat, and Waseem Ahmad Khan, "An Efficient Deep Learning-Based Classifier for an Imbalanced Dataset," MDPI Diagnostics 2022. doi: https://doi.org/10.3390/diagnostics12092115
6. Long Hoang, Suk-Hwan Lee, Eung-Joo Lee, and Ki-Ryong Kwon, "Multiclass Skin Lesion Classification Using a Novel Lightweight Deep Learning Framework for Smart Healthcare" Applied Sciences Journal, 2022. doi:https://doi.org/10.3390/app12052677
7. Shuwei Shen, Mengjuan Xu, Fan Zhang, and Pengfei Shao, "Low-Cost and High-Performance Data Augmentation for Deep Learning-Based Skin Lesion Classification", PMC Article, 2022. doi:10.34133/2022/9765307
8. Anwar Hossain Efat, S. M. Mahedy Hasan, Md. Palash Uddin, and Md. Al Mamun, "Multi-Level Ensemble Framework for Skin Lesion Classification Transfer Learning with Triple Attention" Research Publication, 2024. doi:https://doi.org/10.1371/journal.pone.0309430
9. Md. Shakib Khan, KaziNabiulAlam, AbdurRabDhruba, HasibZunair, and Nabeel Mohammed, "Knowledge Distillation Approaches for Melanoma Detection," Elsevier Publication, 2022. doi:https://doi.org/10.1016/j.compbiomed.2022.105581
10. Muhammad Mateen, Shaukat Hayat, Fizzah Arshad, Yeong-HyeonGu, and Mugahed A. Al-Antari, "Hybrid Deep Learning Framework for Melanoma Diagnosis Demoscopic Medical Images," PMC Publication, 2024. doi: https://doi.org/10.3390/diagnostics14192242
11. Muhammad Zia Ur Rehman, Fawad Ahmed, and Suliman A. Alsuhibany, "Classification of Skin Cancer Lesion Using Explainable Deep Learning", MDPI Sensors Publication, 2022. doi:https://doi.org/10.3390/s22186915
12. Nada M. Rashad, Noha M.M. Abdelnapi, Ahmed F. Seddik, and M. A. Sayedelahl, "Automating Skin Cancer Screening: a deep learning," JEAS / SpringerOpen Publication, 2025. doi:https://doi.org/10.1186/s44147-024-00573-w
13. Saksham Thakur and Sangeeta Sharma, "Ensemble Fusion of ResNet, EfficientNet, and VGG for Skin Lesion Analysis," 15th ICCCNT Conference Publication, 2024. doi: 10.1109/ICCCNT61001.2024.10726235