Skin Lesion Analysis Using Ensemble Learning: A Machine Learning Perspective

Dipali Ratadiya, Kaku Riya Dharmendra*, Monika Changela
Department of Computer Engineering, B. H. Gardi College of Engineering and Technology, Rajkot, Gujarat

Abstract: *Early and precise detection of skin cancer is vital for improving patient survival, as it remains one of the most common cancers worldwide. This study presents a novel ensemble-based skin lesion classification framework that integrates Convolutional Neural Networks (CNN) with Random Forest and XGBoost classifiers. Utilizing the HAM10000 dataset, the CNN model performs feature extraction from dermoscopic images, while Random Forest and XGBoost execute classification. To enhance performance, a hybrid ensemble strategy combining stacking and weighted averaging is employed, effectively leveraging the strengths of individual models. Experimental results demonstrate that the proposed approach significantly improves detection accuracy compared to traditional single-model methods. The developed system not only aids in reliable and automated skin cancer detection but also contributes to raising public awareness and supporting early diagnosis, thereby reducing tumor incidence and improving patient outcomes.*

Keywords: *Machine Learning, Deep Learning, Convolutional Neural Network (CNN), Random Forest, XGBoost, Classification*

1. INTRODUCTION

The millions of people obtain new skin cancer diagnosis every year, and skin cancer is one of the most common and life-threatening kinds of cancer. An early detection and precise classification of skin lesions is crucial to a successful treatment and better chances of a patient survival. Conventionally, visual, thermoscopic and biopsy-based approaches are adopted in the diagnosis of skin cancer by dermatologists. Nevertheless, such traditional techniques may be costly, time-consuming, and subjective to human error, which is why they can cause false diagnoses. Also, in most parts of the world there are limited numbers of trained doctors who practice dermatology; therefore, it is not easy to diagnose many of the patients on time and accurately.Skin cancer is of different types. Actinic Keratosis (AK) is a pre-cancerous disorder following extended sun exposure in the form of rough and scaly patches of skin exposed to sunlight. It has to be treated before it may turn into Squamous Cell Carcinoma (SCC). The most serious type of skin cancer known as melanoma is characterized by moles that are dark in colour and have jagged edges. Melanotic Nevi (moles) are pigmented non-malignant skin lesions which are brown or black in colour and may be either flat or slightly protruded. Another harmless tumour on the skin is dermatofibroma that originates in the dermis and is shaped in the form of hard reddish-brown nodules.

Fig.1 Images of different classes [1]

The least aggressive but the most common skin cancer is Basal Cell Carcinoma (BCC), which starts in the basal cells of the epidermis and normally appears as pearly, pink or red patches that are ulcerated. Benign Keratosis (Seborrheic Keratosis) is a case where in non-cancerous growths that appear as warts, especially in the older adults, can appear like melanoma and therefore distinction should be made to avoid confusion. Squamous Cell Carcinoma (SCC) is the second most frequent kind of skin cancer but more serious and more aggressive compared to BCC; and it is produced by the keratinocytes; it may metastasize when it is not treated. Vascular Lesions comprise abnormal proliferation of blood vessels that can be red, purple or blue in nature and are usually benign although some of them such as angiosarcomas can be malignant.

Medical research of the recent years focuses more on the idea of developing automated skin cancer detection systems with a strong source of power behind them machine learning (ML). A type of ML algorithms, Convolutional Neural Networks (CNNs), has shown exceptional results in diagnosing medical imagery. The use of CNNs allows them to produce a difference between benign and aggressive lesions with high accuracy since they work with large data volumes of dermatological photos.

2. RELATED WORK

The past couple of years marked a breakthrough in the detection of skin cancer thanks to the breakthrough in the field of the deep learning and hybrid machine learning systems. Lubna Riaz et al. combined CNN and Local Binary Pattern (LBP), which is a joint learning network that showed stunning results in accuracy 98.60 percent (testing) and 97.32 percent (validation). Along the same lines, R. Karthik et al. offered a hybrid model based on Swin Transformer and Dense Group Shuffle Non-Local Attention (DGSNLA) and achieved an outstanding performance with F1-score of 96.64 percent on HAM10000 dataset. A CNN-based interpretable model applied by Krishna Mridha et al. based on Grad-CAM and its variations to visualize the model was used to classify using an Android smart healthcare app with an accuracy of 82 percent. In another attempt, Gururaj et al. proposed DeepSkin, based by using DenseNet169 and exploring ResNet50 with utilizing transfer learning, and validated on HAM10000 dataset which achieved significant results via preprocessing and autoencoder to decoder methods.

A number of studies were made on ensemble models and cross-breeding approaches to enhance performance. Ruchi Mittal suggested DermCDSM, deep learning-based clinical decision support system with ICSO algorithms, which proved better results on the ISIC 2017 dataset. Subhajit Chatterjee and others presented IncepX-Ensemble, an ensembles combination of InceptionV3 and Xception via transfer learning and data augmentation, achieving up to 98 accuracy. On the other hand, Subhayu Ghosh used DCNN, Caps-Net, and ViT as feature extracting models and united them through majority voting with 5 ML models (a number that included SVM and XGBoost) to achieve the accuracy of 91.6% at predictions. The multi-model attentional ensemble of Iftekhar Ahmed combined both ResNet50V2 and MobileNetV2 and EfficientNetV2 with high precision and recall accuracy in detecting vascular lesions (99.08% and 99.49%, respectively).

Other methods demonstrate better classification by achieving refined deep learning pipelines. Ahmed Magdy used pre-trained CNNs such as Alex-Net, VGG, Res-Net, and Dense-Net to optimize them through the dim wolf algorithm to recognize ISIC images which obtained a precision of more than 99%. Finally, Sathvika et al. proposed a bi-pipe model, which deemed an Alex-Net CNN and SVM with a bi-sectional texture feature. The SVM pipeline achieved higher accuracy in the HAM10000 dataset with 98.66 percent as compared to CNN. The two pipelines showed high precision on PAD-UFES-20 dataset, where both Alex-Net CNN and SVM had more than 96%. These literatures remind the implications increase of deep learning fusion methodologies and model ensembles to precise multi-class skin lesion arrangement.

3. METHODOLOGY

The proposed machine learning-based technique for detecting skin cancer demonstrates high feasibility at present because of improved medical imaging technology and enhanced computational capabilities and diagnostic models that use artificial intelligence. Large datasets that are properly annotated provide excellent conditions for model training. By using Dense Net and VGG and adding Random Forest and XGBoost the system reaches high accuracy levels which enables its use for real-life diagnosis support through mobile platforms and clinical assistance. The effective application of machine learning requires direct healthcare supplier understanding of predictions together with compliance with ethical and legal standards and interpretability to minimize biases. Patients need the total use of machine learning for healthcare improvement through multiple technical interpretative components.

3.1 DATASET

The analysis employs HAM10000 (Human Against Machine with 10,000 Training Images) from ISIC International Skin Imaging Collaboration for its skin cancer detection functions. There are 10,015 dermatoscopic pictures expanding across seven skin lesion types aimed toward recognizing benign from malignant conditions. Images exist with metadata attributes which contain image_id as well as dx (diagnosis type), dx_type (diagnosis method), and age, sex, and localization (body part affected).

Fig. 2 Dataset HAM10000 [2]

3.2 PROPOSED MODEL

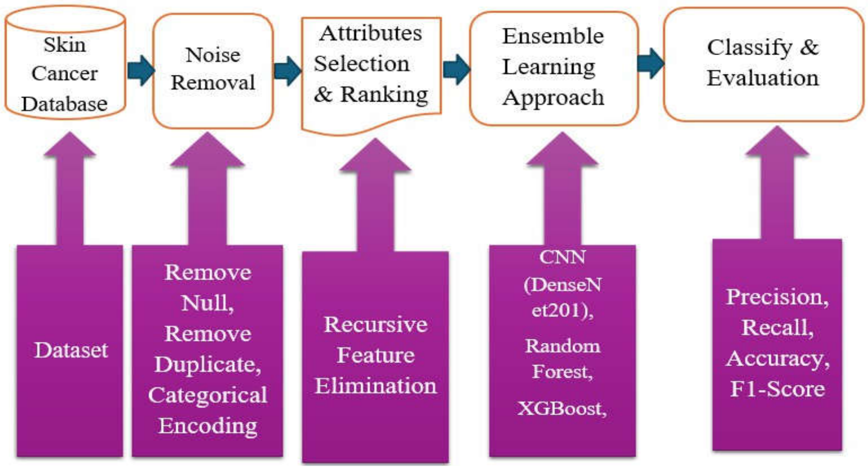

Fig. 3 Proposed model for detection of skin cancer

The steps involved in the detection of skin cancer include loading the set of dermoscopic images like the HAM10000 which has image tags like benign or malignant lesions. Prior to data quality, data preprocessing is done to eliminate the null values and the label encoding to convert the categorical label into numerical space. The process then conducts feature selection with Recursive Feature Elimination that tries to drop less important features in order to achieve more effective classification. A solution based on ensemble learning is chosen, where multiple models are used to achieve greater accuracy, stability, and the overall dependability of a system compared with that of a single-model solution. Important metrics are used to calculate the performance of the classification. Precision indicates the number of correctly identified malignant cases among predicted malignant cases, and recall shows the capacity of the model in identifying all real malignant cases. Accuracy denotes how right the predictions have been. To provide assessment of the performance of the model in a balanced manner, there is the F1-score that balances precision and recall. The measures provide a solid and precise system of skin cancer classification.

3.3 RESULT AND DISCUSSION

The results of skin cancer classification in terms of the performance of different models: they are significant, in particular, each of them provides important data concerning the correctness of diagnoses. DenseNet and Random Forest (RF) gained 93, 93, and 92 percent of accuracy, precision, and recall, respectively. It means that both models were highly reliable in accurately detecting benign and malignant cases with equal sensitivity and specificity. Their confusion matrices would have high true positive and negative rates with little falsities. Although VGG16 performed sufficiently and with reasonable values of 92% and 91% precision and the overall accuracy, respectively, a lower recall (86%) implies that VGG16 was more likely to miss malignant cases- a higher value of false negative records. XGBoost tied with DenseNet and RF (accuracy 93%) and also came in slightly higher in terms of precision 94% which means less false positives in the model but its recall was slightly less at 91%. Lastly, final Model using ensemble had even better performance due to the fact that it had a recall of 92 percent, a precision of 95 percent and an overall accuracy of 94.26 percent. It would show the lowest number of false positives and a good trade-off with false negatives in its confusion matrix and hence it is a very high fidelity and clinically applicable solution of detecting skin cancer.

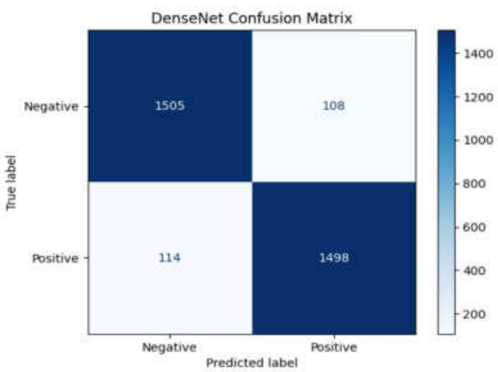

Fig. 4 Des-net

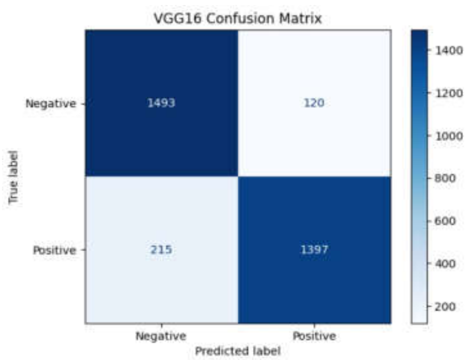

Fig.5 VGG16

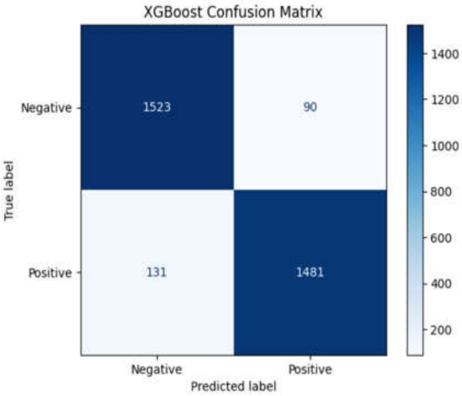

Fig. 6 XGBoost

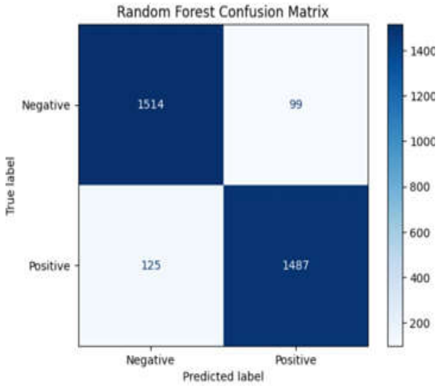

Fig.7 Random Forest

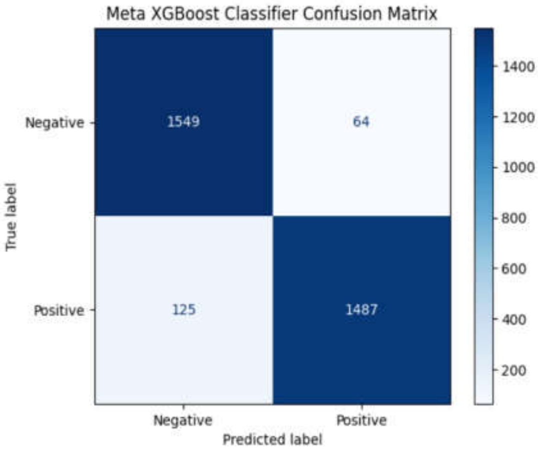

Fig. 8 Meta XGBoost (Final)

The models have been verified on an evaluation dataset shortly after they have all been built. The model can be evaluated utilizing a range of performance metrics, including precision, recall and accuracy. Also, the confusion matrix is used for performance measures.

Table 1. Result with performance measures of All models

Method	Accuracy (%)	Precision (%)	Recall (%)
DenseNet	93	93	92
VGG16	91	92	86
Random Forest	93	93	92
XGBoost	93	94	91
Final Model	**94.26**	**95**	**92**

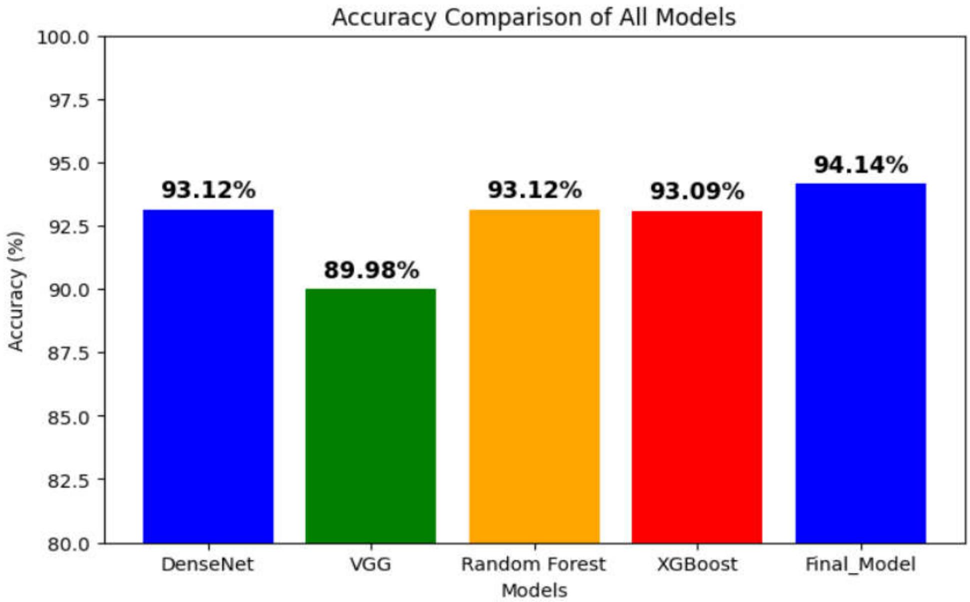

Fig. 9 Shows model Accuracy

A classification model displays its performance through the Receiver Operating Characteristic (ROC) Curve while using different threshold values. The True Positive Rate and False Positive Rate measurements on the ROC Curve are represented through a graphical plot of multiple decision threshold values.

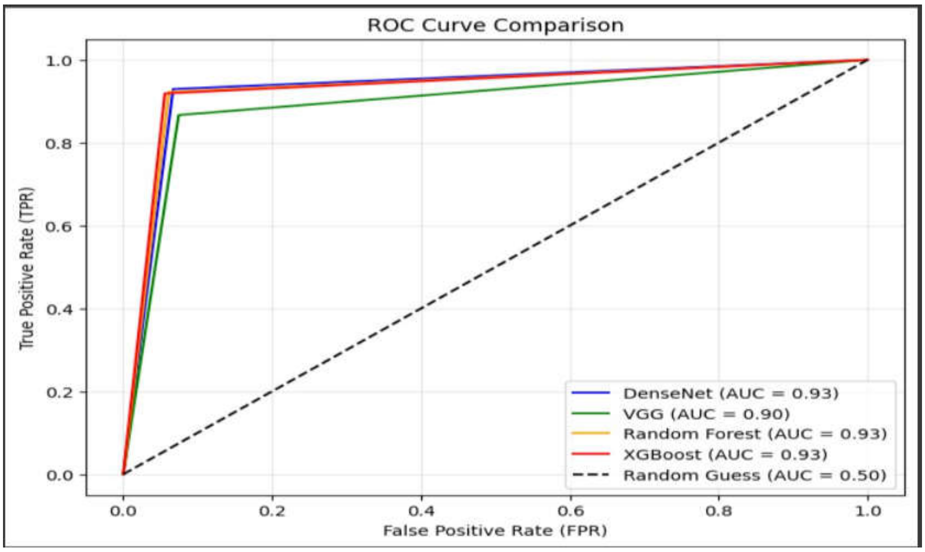

Fig.10 Shows ROC Curve of All Models

4. CONCLUSION

The research develops an accurate machine learning method for skin cancer detection which utilizes CNN models and DenseNet architecture along with ensemble algorithms that include Random Forest and XGBoost. The research shows that DenseNet and VGG among CNN models demonstrate excellent effectiveness in benign versus malignant lesion identification. The implementation of ensemble learning methods increases model performance through classifier combination to gain better robustness and generalization abilities. The DenseNet model reached 93% accuracy as the best result with XGBoost and Random Forest reaching 93% accuracy. The ensemble model using DenseNet with Random Forest and XGBoost achieved a minor enhancement over previous results but kept an accuracy value of 94.11%. The model needs additional enhancements to reach accurate and reliable performance standards suitable for clinical uses.

REFERENCES

1. Riaz, L., Qadir, H.M., Ali, G., Ali, M., Raza, M.A., Jurcut, A.D., Ali, J.: A comprehensive joint learning system to detect skin cancer. IEEE Access (2023).
2. Karthik, R., Menaka, R., Atre, S., Cho, J., Easwaramoorthy, S.V.: A Hybrid Deep Learning Approach for Skin Cancer Classification using Swin Transformer and Dense Group Shuffle Non-Local Attention Network. IEEE Access (2024).
3. Mridha, K., Uddin, M.M., Shin, J., Khadka, S., Mridha, M.F.: An interpretable skin cancer classification using optimized convolutional neural network for a smart healthcare system. IEEE Access 11, 41003–41018 (2023).
4. Gururaj, H.L., Manju, N., Nagarjun, A., Aradhya, V.M., Flammini, F.: DeepSkin: a deep learning approach for skin cancer classification. IEEE Access 11, 50205–50214 (2023).
5. Mittal, R., Jeribi, F., Martin, R.J., Malik, V., Menachery, S.J., Singh, J.: DermCDSM: clinical decision support model for dermatosis using systematic approaches of machine learning and deep learning. IEEE Access (2024).
6. Chatterjee, S., Gil, J.M., Byun, Y.C.: Early Detection of Multiclass Skin Lesions Using Transfer Learning-based IncepX-Ensemble Model. IEEE Access (2024).
7. Ghosh, S., Dhar, S., Yoddha, R., Kumar, S., Thakur, A.K., Jana, N.D.: Melanoma Skin Cancer Detection Using Ensemble of Machine Learning Models Considering Deep Feature Embeddings. Procedia Computer Science 235, 3007–3015 (2024).
8. Ahmed, I., Routh, B.B., Kohinoor, M.S.R., Sakib, S., Rahman, M.M., Azzedin, F.: Multi-Model Attentional Fusion Ensemble for Accurate Skin Cancer Classification. IEEE Access (2024).
9. Magdy, A., Hussein, H., Abdel-Kader, R.F., Abd El Salam, K.: Performance Enhancement of Skin Cancer Classification Using Computer Vision. IEEE Access (2023).
10. Google: Mathematics, https://www.mdpi.com/2227-7390/12/7/1030, last accessed 2025/09/15.
11. Naukri Campus: Types of Machine Learning, https://www.naukri.com/campus/career-guidance/wp-content/uploads/2024/04/types-of-machine-learning.png, last accessed 2025/09/15.
12. Analytics Vidhya: Convolutional Neural Network Bird Example, https://editor.analyticsvidhya.com/uploads/804084200125366Convolutional_Neural_Network_to_identify_the_image_of_a_bird.png, last accessed 2025/09/15.
13. MDPI: Cancers Journal Image, https://www.mdpi.com/2072-6694/15/4/1183, last accessed 2025/09/15.

14. MDPI: Diagnostics Image, https://www.mdpi.com/diagnostics/diagnostics-12-00726/article_deploy/html/images/diagnostics-12-00726-g001.png, last accessed 2025/09/15.

Author contributions

Methodology, software, analysis, investigation, data curation, visualization, and original draft, Dipali Ratadiya; Conceptualization, supervision, validation, review & editing, Kaku Riya Dharmendra and Monika Changela.

Acknowledgments

This research received no external funding.