

Analysis Of Skin Lesion With Ensemble Learning Classifiers And Image Denoising

Anup Lal Yadav¹, Preeti Rana², Santosh Kumar³, Soumya G D⁴, Tejbir Singh^{5*}, Nithya B A⁶

^{1,2}Computer Science and Engineering, School of Engineering and Technology, CGC University Mohali, Punjab, India 140307,

³Department of CSE, Chandigarh University Mohali, Punjab, India

⁴School of Computer Science and Engineering, Presidency University Bengaluru, India

⁵M M Institute of Computer Technology & Business Management, Maharishi Markandeshwar (Deemed to be University), Haryana, India

⁶Presidency School of Information Science, Presidency University Bengaluru, India

¹anupsaran@gmail.com, ²ranapreeti106@gmail.com, ³santosh.iete@gmail.com,

⁴soumya@presidencyuniversity.in, ^{5*}tejbirrana662@gmail.com@gmail.com,

⁶nithya.ba@presidencyuniversity.in

*Corresponding Author

Abstract: The screening of skin cancer at a tender age and at the correct time is the only way of ensuring that more patients survive cancer as skin cancer remains one of the most common forms of malignancies in the world today. Computerized skin lesion analysis has been of much interest as it can help dermatologists in the early diagnosis of lesions and minimize the errors of subjectivity in their interpretation. This paper is a new ensemble-based skin lesion classification model that combines both deep learning and machine learning to ensure effective and dependable diagnostic accuracy. A Convolutional Neural Network (CNN) is used in the proposed method to extract features of dermoscopic images according to their complex space and texture features of skin lesions automatically. Deep feature is then classified with the help of Random Forest and Extreme Gradient Boosting (XGBoost) classifiers. To even increase the predictive accuracy and generalization power, a hybrid ensemble approach based on stacking and weighted averaging are proposed, which can conveniently make use of the complementary strengths of models. The framework is tested on the publicly available HAM10000 dataset that consists of various categories of pigmented skin lesions of different visual appearances. A long series of experimental results show that the suggested ensemble model has a dramatically positive advantage in performance regarding accuracy, sensitivity, specificity as well as general reliability over the individual classifiers and traditional single-model practices. The findings help to emphasize the success of the integration of deep feature learning and ensemble decision-making. The acquired system is a valid and automatic instrument in detecting skin cancer in order to aid the process of early diagnosis, clinical decision-making, and these factors may be used to reduce the development of tumors and increase the survival rates of patients.

Keywords: Skin Lesion, Ensemble Classifier, Deep Learning, CNN, Random Forest, XGBoost, Classification.

I. Introduction

The millions of individuals receive new skin cancer diagnosis annually, and one of the most frequent and life-threatening types of cancer is skin cancer. A successful treatment and higher chances of a patient survival should be achieved through early detection and accurate classification of skin lesions. Traditionally, dermatologists consider adopting visual, thermoscopic and biopsy-based methods to diagnose skin cancer. However, traditional methods can be very expensive, time-consuming and prone to human error and, thus, lead to false diagnosis. Moreover, trained dermatologists are few in most regions of the world, hence there is no ease in diagnosing most of the patients early and correctly. Skin cancer is of various types. Actinic Keratosis (AK) is a non-cancerous disease that is caused by a prolonged exposure of the skin to the sun in the shape of rough and scaly dots

of skin surface. Before it develops into Squamous Cell Carcinoma (SCC), it must be treated. The worst form of skin cancer is a skin cancer referred to as melanoma, which is dark coloured with jagged edges. The moles are Melanotic Nevi (pigmented non-malignant lesions of the skin) that are coloured brown or black and can be flat or slightly raised. Dermatofibroma is another harmless skin tumour that develops in dermis and is formed in shape of hard reddish-brown nodules. One of the most prevalent and fastest-growing types of cancer across the globe is skin cancer that is a significant problem to the global public health in the developed and developing world. As it is reported by the international statistics on cancer, skin-related malignancies have been increasing gradually over the last several decades, because of the long-time exposure to ultraviolet (UV) radiation, environmental pollution, lifestyle modifications, as well as genetic inclination [1]. Categories encompass skin cancer inclusive of non-melanoma and melanoma with melanoma being the most aggressive and dangerous type of cancer. Though melanoma is less prevalent among all the cases of skin cancer, it contributes most of the death rates of skin cancer as it has the highest propensity of metastasis. Conversely, non-melanoma skin cancers like basal cell carcinoma and squamous cell carcinoma are more common but are less aggressive and with better prognosis. Being able to detect malignant lesions in the skin at an early stage leads to a high survival rate, where research has shown that the early diagnosis of Melanoma at an early stage has a high five-year survival rate of over 95 percent. Skin lesions analysis should therefore be conducted accurately and timely, to make adequate clinical decisions and manage patients.

The main mode of skin lesion analysis is the visual inspection of the suspicious areas on the skin surface with the aid of dermoscopy features. Dermoscopy is a technique that allows dermatologists to view the structures of the skin under the surface that cannot be seen with the naked eye, which is more accurate in diagnosing compared to traditional clinical examination. Nevertheless, dermoscopic analysis of digital images is one of the most subjective and is heavily reliant on experience, skills and workload of the attending clinician. Inter-observer variation may be very high even with trained dermatologists and when distinguishing between benign lesions like nevi and malignant melanoma. Such ambiguity of diagnosis predisposes misclassification, late treatment, biopsies that are unnecessary. With the constantly increasing number of patients and the already low level of resources in healthcare, manual examination can simply not be relied on as a consistent and scalable method. These issues have led to an increased pace in the creation of automated and semi-automated computational means to conduct the analysis of skin lesion. The computer-aided diagnosis (CAD) systems have become potent to aid dermatologists in identifying and categorizing skin lesions with greater accuracy and efficiency. The skin lesion analysis with a typical CAD system consists of several processing steps and stages, such as image acquisition, preprocessing, lesion segmentation, feature extraction, and classification. Preprocessing is one of these steps that are very important to enhance the quality of images and the qualities of discriminative lesions before further analysis [2]. The dermoscopic images tend to undergo several forms of degradation that are brought about when capturing the images which include sensor noise, uneven illumination, motion blur, low contrast, and compression artifacts. Such distortions seriously deteriorate the visual quality and cause blurring of clinically significant features thus restricting the outcome of the automated diagnostic systems. Thus, proper preprocessing schemes are impossible to do without, to achieve healthy and sound lesion classification.



Fig.1 Images of different classes [1]

Noise is one of the most notable problems of a dermoscopic image analysis. There are various sources of noise, such as imaging sensors, the state of ambient lighting, patient motion, the change in dermoscopes devices. Random noise also decreases the clarity of the image and disturbs the fine texture patterns on which the characterization of the lesion is required. Also, variation of illumination is another issue that exists in dermoscopic datasets. Abnormal lighting will give shadow areas, white reflections, and uneven distribution of color in the pictures. These differences give a false perception of the boundaries of the lesion and changes the intensity of colors and these are important parameters in dermatological examination. The illumination anomalies will confuse a segmentation algorithm without its correction and have a negative impact on feature extraction. The weak contrast of the lesion areas and normal skin also makes it difficult to detect the lesions. The color and texture variations of cancerous lesions and normal skin are very inconspicuous in most cases; thus boundary demarcation is very challenging. This problem especially comes out in the initial stages of melanoma where the visual difference is very slight. Additionally, dermoscopic pictures normally have artifacts including hair strands, air bubbles, ruler marks, ink marks and blood vessels [3]. These artifacts blur musculoskeletal lesion regions and may give false structural images that can be wrongly interpreted as lesion. One of the most dominating sources of interference is hair occlusion and this has a great impact on the segmentation accuracy and later classification performance. Image denoising and the removal of artifacts must be considered as critical elements of the current skin lesion analysis systems to overcome these difficulties. Image denoising is the task of removing undesired noise and retaining the valuable structural information in form of edges, texture, and colour variation. The most common traditional filtering methods like Gaussian filtering, median filtering and bilateral filtering have been extensively used to eliminate noise. There is, however, a possibility of excessive smoothing, which can obscure the edges of lesions, and eliminate distinguishing characteristics. More recent denoising methods such as non-local means filtering and wavelet-based denoising offer better noise reduction due to the use of spatial redundancy and multi-resolution analysis [4]. Combination of efficient denoising methods results in better visualization and lesion segmentation accuracy and eventually leads to better classification performance. Ensemble learning has become a viable alternative and complementary approach to enhance the classification performance and the generalization capability. Ensemble learning is a method that engages the outcomes of a

series of base classifiers to come up with better accuracy than when using single models. The basic rule of ensemble learning is the variance reduction, minimizing bias and increasing noise and overfitting resistance. Bagging, boosting, random forests, voting classifiers, and stacking ensembles are some of the techniques that have shown good performance in a variety of medical diagnostic tasks. Ensemble models are found to be effective in the analysis of skin lesions, where various classifiers complement each other and hence improving the ability to distinguish between benign and malignant lesions.

The suggested research is dedicated to the consideration of efficiency of image denoising methods together with ensemble of learning classifiers that provide correct diagnosis of skin lesions. Several denoising methods are explored to find out their effects on visual quality and classification. After preprocessing, clinically relevant characteristics of color, texture, and shape are obtained to reflect lesion characteristics in detail. Multiple machine learning classifiers are also trained and pooled together by employing ensemble approaches and voting and stacking. The framework is strongly tested on a basis of standard dermoscopic datasets and achieves reliable performance results based on established performance measures. The most important findings of this study are as follows. To begin with, an elaborate preprocessing pipeline is created to deal with noise, variation of illumination, low contrast, and artifact interference of dermoscopic images. Second, several image denoising algorithms are systematically explored and evaluated by quantitative measures of quality to find the most efficient strategy of reducing noise [5]. Third, a powerful ensemble learning system is drawn through the combination of several base classifiers to strengthen the classification. Fourth, the effect of image denoising on the classification performance is properly investigated based on the experimental assessment. Lastly, it has been shown that the offered method has a higher diagnostics accuracy in comparison with single-classifier models and other currently existing baseline methods, which indicates its possible practical implementation in actual clinical settings.

The rest of the paper is structured in the following way. Section 2 is a review of related work in skin lesion analysis, image denoising and the ensemble learning-based classification. Section 3 focused on the image preprocessing and denoising methods with the proposed methodology and elaborates the ensemble learning classifiers models. Section 4 covers the description of the experimental arrangement and results. Section 5 is provided where key findings and contributions are presented.

II. Literature Review

The previous two years saw the discovery of the breakthrough in detecting skin cancer due to the breakthrough in the sphere of the deep learning and hybrid machine learning systems. The CNN and Local Binary Pattern (LBP) is a joint learning network, and it was used in combination with CNN, and in this case, the combination resulted in an amazing accuracy of 98.60 percent (testing) and 97.32 percent (validation), as demonstrated by Lubna Raza et al. Similarly, authors [6] provided a hybrid model that was built on Swin Transformer and Dense Group Shuffle Non-Local Attention (DGSNLA) and scored exceptionally with F1-score of 96.64 percent on HAM10000 dataset. Researchers [7] used an Android smart healthcare app with an accuracy of 82 percent to classify with a CNN-based interpretable model applied using Grad-CAM and various variants to visualize the model. In yet another effort, Gururaj et al. came up with DeepSkin, founded based on DenseNet169 and experimenting on ResNet50 with transfer learning use, and tested on HAM10000 dataset that yielded impressive results using preprocessing and autoencoder to decoder architecture. Ensemble models and crossbreeding methods were made in several studies to improve performance. Ruchi Mittal proposed DermCDSM, deep learning-based clinical decision support system with ICSO algorithms that had a superior

performance on the ISIC 2017 dataset. IncepX-Ensemble, a combination of Inception V3 and Xception using transfer learning and data augmentation was introduced by [8] and others with up to 98 accuracies. Authors [9], on the other hand, employed DCNN, Caps-Net and ViT as feature extracting models and combined them together using majority voting with 5 ML models (which also included SVM and XGBoost) to reach the accuracy of 91.6% at predictions. The multi-model attentional ensemble of [10] used both ResNet50V2 and MobileNetV2 and EfficientNetV2 with high accuracy and recall rate in the detection of vascular lesions (99.08% and 99.49% respectively). Other techniques show more refined classification through the accomplishment of the deep learning pipelines. [11] employed the pre-trained CNNs like Alex-Net, VGG, Res-Net, and Dense-Net to tune them using the dim wolf algorithm to identify ISIC images that achieved a precision of over 99%. Lastly, Sathvika et al. came up with a bi-pipe model that considered an Alex-Net CNN and SVM with a bi-sectional texture feature. The SVM pipeline was more accurate in the HAM10000 data with a score of 98.66 percent in comparison to CNN. On PAD-UFES-20 data, both Alex-Net CNN and SVM were more than 96% precise on the two pipelines. The implications increase is reminding the use of deep learning fusion methodologies and model ensembles to the exact arrangement of the skin lesion of multiple classes.

The classification of skin lesions as a field of study has also been a field of research of over 20 years since skin cancer has been on the rise and the need to have an efficient automated diagnostic tool. The first research mostly used conventional machine learning methods, which were based on manual feature extractors and traditional classifiers. These methods were mostly sequential pipeline with image preprocessing, lesion segmentation, feature extraction, and classification. Features on color like mean intensity, variance and histogram distributions were commonly applied because of the high correlation between lesion pigmentation and malignancy. There were also several studies done on texture descriptors such as Gray Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP), Gabor filters, among others, aimed at exploiting the structural irregularities of the lesion areas. The shape-based features, including asymmetry, irregularity of the border and diameter (usually based on the ABCD dermatological rule), were deemed important in the differentiation between benign and malign lesions. These artisanal characteristics were then categorized with machine learning models like the support vector machines, k nearest neighbor, decision trees, random forest and naive bayes classifiers. Despite these methods showing reasonable results on small and medium-sized datasets, their results were very dependent upon the quality of feature engineering and precision of segmentation. Also, handcrafted features were not strong to illumination change, noise, or high intra-class variations that are observed in dermoscopic images [12].

Deep learning-based lesion analysis methods have been of great interest with the development of computational power and large and annotated datasets. Convolutional neural networks (CNNs) have proved to be quite successful in medical image classification as they directly learn hierarchical features representations by looking at raw images. Various research used pre-trained deep architectures including VGGNet, ResNet, DenseNet, Inception and EfficientNet with the transfer learning strategies to address the small dataset volume. Such models have not only performed at the same or better classification performance than expert dermatologists in controlled experimental conditions but have surpassed them by a margin. More complex architectures have combined attention, multi-scale feature extractors, and CNN-RNN structures to achieve even better lesion discrimination [13]. Besides that, U-Net and its variants have been used to localize lesion areas prior to classification. Although deep learning methods are highly accurate, they have several weaknesses such as, they are computationally complex, require large memory, are sensitive to hyperparameters, and lack interpretability. Massive, labelled datasets are also required, and the fact that they can experience domain

shift across imaging devices also limits their application in resource constrained clinical settings. Medical image analysis heavily relies on image preprocessing and denoising tools because the raw clinical images usually have a lot of noise and artifacts that deteriorate the diagnostic accuracy. Noise in dermoscopic imaging is caused by sensor defects, variations in light, and patient motion whereas artifacts, including hair, air bubbles, ruler marks, etc, further distort lesion architecture. The preprocessing methods used have been traditional preprocessing methods like median filtering and Gaussian smoothing which have been widely used to reduce noise. Median filtering works well especially in dealing with impulse noise, but a smooth transition of the intensity is available with Gaussian filtering. Nevertheless, the techniques can result in blurriness of fine edges and boundaries of the lesions in the event of excess. To reduce this shortcoming, the superior degree of denoising methods such as bilateral filtering and non-local means filtering has been presented to retain the edge information whilst inhibiting noises. Wavelet-based denoising techniques have also become popular because it is able to analyse images on various levels of resolution and pick out noise components. Histogram equalization and contrast-limited adaptive histogram equalization (CLAHE) are contrast enhancement methods that are commonly used to enhance the visibility of lesions. Although preprocessing techniques have been considered important, in many cases they are considered as supplementary measures and little has been done to explicitly analyze the direct influence of preprocessing techniques on classification accuracy [14].

To enhance the reliability of prediction and generalization outcomes, ensemble learning has become one of the strong paradigms in medical diagnosis systems. Ensemble classifiers group together several base classifiers to take advantage of their respective strong points, so that the risk of overfitting is lowered and their resistance to both noise and data imbalance is enhanced. Random forests, which are bagging based methods, can also improve stability whereby many decision trees trained on bootstrap samples can be aggregated. The methods, such as AdaBoost and Gradient Boosting, are aimed at the enhancement of weak classifiers through the priority given to the cases of misclassification. Voting based ensembles combine heterogeneous classifiers with majority or weighted decision rules and stacking ensembles combine the output of base models to the best of their output. Ensemble learning has also shown better results in medical imaging applications like breast cancer, brain tumor, retinal disease and skin lesion applications. Multicast models are especially suitable in classification under dermoscopic images because of the similarity between the inter-class and intra-class variation, which pose problems in the learning of single models. Although significant advancements have been made based on conventional machine learning, deep learning and ensemble-based models, there are still a number of limitations existing in the literature. Several machine learning-based methods are sensitive to lesion segmentation accuracy, and segmentation inaccuracies considerably propagate into the classification system. Despite being exceptionally precise, deep learning techniques frequently serve as black-box systems with minimal explainability, making them less acceptable by clinicians. Moreover, the majority of the deep learning models require a huge amount of computer power and cannot be implemented in real-time or mobile diagnostics. The other crucial problem is the lack of balance of datasets because the number of malignant cases is much lower compared to benign ones, which predisposes biased predictions. Separate studies are concerned much with optimization of the classifiers but ignore effect of degradation of image quality due to noise and artifact. Most of the available literature uses preprocessing heuristically without making comparative analyses of denoising methods. With the assistance of the literature analysis, it is possible to identify a research gap. Little research on the combined performance of image denoising and ensemble learning on the performance of the skin lesion classification task is conducted. Most studies perceive

preprocessing and classification as separate entities as opposed to being modules that form a single unit. Also, the relative comparison of several denoising methods with the classification accuracy is still lacking. A systematic method to assess the effect of reduction of noise on the quality of features and the ensemble decision-making process is required. To fill these gaps, the current research will offer a combined framework that entails the use of optimized image denoising methods with ensemble learning classifiers to increase robustness, reliability, and accuracy of diagnosis in automated skin lesion detection.

III. Proposed Methodology

The suggested methodology introduces a combined approach to automated lesion analysis that would integrate a powerful image preprocessing, efficient lesion segmentation, discriminative feature extraction, and classification with the help of the ensemble learning. The key goal of the suggested solution is to increase the diagnostic reliability by increasing the image quality with the help of the sophisticated denoising methods and improving the classification performance with the ensemble learning classifiers. The methodology has three principal sub-sections; description and preprocessing of the dataset, lesion segmentation and feature extraction, and ensemble learning based classification structure. The general flow of the proposed system is that dermoscopic images are first acquired, then pre-processed and any artifact removed, the lesion area is then segmented, clinically relevant features are extracted and finally the lesion is classified by an optimized ensemble model. The processes that entail the process of skin cancer detection are loading the set of dermoscopic images such as the HAM10000 that contains image marks such as benign or malignant lesions. Before data quality, there is the process of data preprocessing to remove the null data and the label encoding of the categorical label into numerical space. It then performs feature selection using Recursive Feature Elimination which attempts to remove the less relevant features to get a more effective classification. An ensemble learning-based solution is selected, which involves multiple models to provide more accuracy, stability, and general reliability of a solution than a solution based on a single model. The performance of the classification is calculated with the help of important metrics. Precision and recall denote the number of malignant cases recognized correctly among the predicted malignant cases and the ability of the model to identify all the real malignant cases. Accuracy is used to refer to the extent to which the predictions have been correct. To offer an evaluation of the performance of the model in a balanced fashion, there is the F1-score that balances both the precision and the recall. The measures offer a secure and accurate regime of skin cancer classification.

3.1 Dataset Description and Image Preprocessing

To have the strength and generalizability of the proposed framework, publicly available benchmark dermoscopic image datasets were used. In the given study, the HAM10000 data was chosen because it is widely used in the research community and its extensive coverage of pigmented skin lesions. The HAM10000 database consists of 10,015 dermoscopic images that were obtained in various clinical sources with the help of various acquisition instruments. These records consist of seven different lesion types, i.e. melanoma (MEL), melanocytic nevi (NV), basal cell carcinoma (BCC), actinic keratoses (AKIEC), benign keratosis-like lesions (BKL), dermatofibroma (DF), and vascular lesions (VASC). The images are given with different resolutions and in most case at 450 x 600 to 600 x 800 pixels thus bringing about inconsistency in the dimensions and color distribution. The issue of extreme class imbalance is one of the key problems with the HAM10000 dataset. The dataset is dominated by benign lesions (melanocytic nevi) and has a relatively smaller number of samples of malignant lesions (melanoma and dermatofibroma, etc.). This imbalance may cause classification models to be biased with

majority classes and decrease sensitivity to important cases of malignancies. The stratified sampling was used to reduce this problem by ensuring that the class distribution in training, validation, and testing subsets was maintained. The dataset was separated into 70 training data, 15 validation data and 15 testing data. Hyperparameter optimization and model selection were done on the validation set and the testing set was not seen at all during the training since unbiased evaluation of the model performance would have been required.



Fig. 2 Dataset HAM10000 [2]

Before extracting features and classifying them, all dermoscopic images were preprocessed using a lengthy pipeline. Images were acquired and resized and normalized, to bring the spatial dimensions to similar levels across the dataset. All images had been scaled to a constant resolution size of 256 x 256 pixels so that they all represented similar inputs and so that computation would be made easier. The pixel intensity was normalized to reduce the differences that are caused by the various imaging devices. As the color information is important in the diagnosis of the skin lesions, color space transformation was included during preprocessing stage. The RGB color space was transformed to other color spaces including HSV color space, LAB color space. The HSV color model divides chromatic data and data on intensity, making pigmentation patterns more easily differentiated, whereas the LAB color space provides perceptual consistency, which increases contrast between lesion and surrounding skin. Image denoising was an essential part of preprocessing phase. Gaussian noise, salt-and-pepper noise and low-frequency illumination artifacts tend to affect dermoscopic images. Several denoising algorithms were explored to find the best noise cancelling algorithm. To dispell impulse noise but retain edges, median filtering was used. Intensity variations were smoothed with the aid of Gaussian filtering though it can cause slight blurring. Edges preserving smoothing was done using bilateral filtering that considers both the spatial and intensity proximity. To use repetition patterns in images, the non-local means denoising was applied to take advantage of them and provide a better noise reduction without losing texture details. Also, wavelet-based denoising was carried out by breaking down images into multi-resolution frequency components and selective removal of coefficients that were mostly dominated by noise.

Image quality was further improved by removing artifacts. One of the most prevalent dermoscopic artifacts such as hair occlusion was eliminated with the DullRazor algorithm. This method identifies structures of hair with morphological applications and fills damaged pixels with bilinear interpolation. Contrast-limited adaptive histogram equalization (CLAHE) was employed to better visualize the lesions, and it maximizes the local contrast without exaggerating noise. Peak signal to noise ratio (PSNR), mean squared error (MSE) and structural similarity index measure (SSIM) were quantitatively used to

determine the effectiveness of each denoising technique. Such metrics have given objective evaluation of the quality of images and helped in choosing the best preprocessing measures that could then be used in analysis.

3.2 Lesion Segmentation and Feature Extraction

Lesion segmentation is an important process because it separates the area of interest of the lesion to healthy skin and other artifacts (and other skin lesions). Accurate segmentation has a direct impact on the accuracy of features extracted and the results of classification. Different segmentation methods were tested in the proposed methodology to determine the best method of segmentation to use when identifying the boundaries of lesions. To begin with, Otsu thresholding procedure was used to classify the foreground and background using the maximum variance between the classes. Despite its computational simplicity, thresholding algorithms are frequently not able to cope with illumination variations and complicated lesion architecture. In order to work around these weaknesses, K-means clustering was employed to divide image pixels into groups according to color similarity. The area of lesion was defined by picking groups of dominant pigments. Active contour models were also used to sharpen the edges of the lesion by minimizing energy to allow the smooth evolution of contours to the edges of objects. To improve further, a hybrid form of segmentation that uses U-Net convolutional neural network was used as an option. U-Net architecture showed better segmentation accuracy because it learned spatial context and boundary feature especially in the difficult images that have low contrast and irregular borders. After segmentation, morphological post-processing like erosion, dilation, filling of holes and boundary smoothing were used to remove spurious areas and clean up the lesion contours. Standard measures of overlap were used to measure the segmentation performance, such as the Dice similarity coefficient and the Jaccard index. These measures are used to measure consistency between the predicted mask of the segmentations and ground-truth labels, which are good indicators of segmentation accuracy.

After the accurate segmentation of the lesion area, the extraction of the features was conducted to present the lesion characteristics numerically. Three features, namely color, texture, and shape features, were obtained. Variables in color were statistical values of mean, variance, skewness calculated in the various channels of colors. Histogram features were identified to reflect pigmentation patterns of distribution that is mostly related to melanoma progression. The extraction of texture features was carried out using Gray Level Co-occurrence Matrix (GLCM) that provided the second-order statistical measures of contrast, correlation, energy, and homogeneity. The Local Binary Patterns (LBP) were used to encode local texture changes and micro-patterns in the regions of lesions. Orientation- and frequency-specific texture responses were extracted using Gabor filters, which made it possible to represent texture multi-scale. The shape features were obtained to measure the geometry of the lesions and morphological aberrations. Parameters of lesion area, perimeter, compactness, and eccentricity were calculated. The evaluation of asymmetry was done by comparing half of the lesions along the orthogonal axes whereas the assessment of border irregularity was done using the contour complexity measures. The estimation of the diameter was made according to the extent of the maximum lesions. Such shape descriptors are consistent with the clinically defined ABCD dermatology rule and make medical interpretation of the extracted features more interpretable. All the extracted attributes were then concatenated to form the final feature vector and then feature normalization was done to achieve constant scaling before classification.

3.3 Ensemble Learning-Based Classification Framework

The last step in the proposed methodology is devoted to classification with the help of ensemble learning methods. Due to the great variability of dermoscopic images and the commonality of features between lesion types, use of a single classifier in most cases would lead to wobbly predictions. To overcome this weakness, various base classifiers were used, and they are SVM, RF, k-nearest neighbors (KNN), DT and naive bayes (NB). All classifiers have distinct learning patterns SVM has good generalization behavior based on the maximization of the margin, RF has strengths in terms of robustness due to its use of randomized decision trees, KNN displays instance similarity, DT provides interpretability, and NB has probabilistic classification effectiveness. Ensemble techniques were introduced to enhance predictive performance further. Ensembles that used bagging were used to minimize variance by training numerous classifiers using bootstrap samples. AdaBoost and Gradient Boosting were boosting algorithms that were used to enhance weak learners after repeated use to concentrate on misclassified data. Both the soft and hard voting were used in constructing the voting classifiers. Hard voting implies majority-based classification of classes, and soft voting implies average-based classification of classes, which provides better stability. A stacking ensemble model was also generated in which base classifier predictions were used as inputs by a meta-classifier. Such an architecture allowed the adaptive weighting of the outputs of the classifier thus maximising the final decision. The grid search and cross-validation were used to hyperparameter tune and determine the best model settings. The validation dataset was used to optimize parameters like the type of kernel and regularization constant to use in SVM, number of trees to use in random forest, value of k in KNN as well as the learning rates to use in the boosting algorithms.



Fig. 3 Proposed model for detection of skin cancer

The suggested ensemble structure combines quality assessment of preprocessing, discriminative feature learning, and the diversity of classifiers in a single system. The architecture also guarantees a greater resilience to noise, class imbalance and feature variability. The proposed methodology gives high diagnostic performance when compared with the traditional method of using single-model image denoising methods, thus making it a dependable and scalable technique to automated analysis of skin lesions, through the implementation of ensemble learning classifiers.

The proposed machine learning-based machine to detect skin cancer is very feasible at this point of time due to the advancement of medical imaging and higher computational power and diagnostic models that employ artificial intelligence. Huge datasets that are annotated appropriately present superior environments to train a model. Dense Net and VGG with the addition of Random Forest and XGBoost allow the system to reach high rates of accuracy that make it possible to use it to support the diagnosis process in real-life by means of mobile devices and clinical support. To ensure that machine learning is effectively implemented, it is necessary that the supplier of direct healthcare has firsthand

knowledge of the predictions and adheres to ethical and legal standards and interpretability to reduce biases. Patients require the full application of machine learning in healthcare enhancement using various technical interpretative elements.

IV. Experimental Setup and Results

A standardized hardware and software environment was used to test the proposed skin lesion analysis framework experimentally to be able to reproduce and draw a fair comparison. All the experiments were conducted in a workstation with an Intel Core i7 processor, 16 GB RAM and an NVIDIA graphics card with 6 GB memory. Python programming language with popular libraries such as OpenCV to process images, NumPy and SciPy to perform numerical calculations, Scikit-learn to perform machine learning and ensemble modeling, and Matplotlib to visualize the data were used to implement the software. Experiments were carried out in an operating system environment on Windows/ Linux. Tuning of the parameters was vital in attaining good performance of the system. In the case of support vector machine classifier, radial basis function kernel was used with regularization and grid search based on optimization of kernel parameters. To balance between accuracy and computational efficiency, the random forest classifier was set with 100-300 decision trees. The k nearest neighbors classifier was used and the value of k was chosen through empirical means. The depth of the decision trees was limited to curb overfitting and the naïve Bayes used Gaussian probability distribution. To enhance the algorithms, both the learning rate and estimators were properly tuned on validation data. The selection of all ensemble parameters was made according to a validation performance in order not to leak information provided by the test dataset.

An approach of stratified k-fold cross-validation was implemented to perform strong performance estimation in the conditions of class imbalance. The data was separated into five folds, and the proportion of classes was maintained in each fold. In the training process, the learning and testing folds were 4 and 1 respectively and the process was repeated until every fold was used as the test set one time. The average calculated results were the final performance results. Such an approach reduced the bias introduced by partitioning of the data and increased the reliability of generalization. The Medical diagnostic measures that are universally accepted were used to measure the performance of the suggested framework. Overall classification correctness was used to measure accuracy, whereas the reliability of positive predictions was measured by precision. Recall also known as sensitivity tested the capability to correctly detect malignant lesions. Specificity acquired the measurement of correct identification of benign cases and that is paramount to minimize unnecessary biopsies. The F1-score offered a balanced score in terms of recall and precision. Also, the receiver operating characteristic (ROC) curve and area under the curve (AUC) was utilized to compare the discriminative capability at various decision thresholds.

Image denoising methods were tested in terms of visual and quantitative performance. Visual evaluation showed that median filtering was useful to remove impulse noise but with slight edge blurring. The smoother images were made with the Gaussian filtering, but sharp edges were diminished. Bilateral filtering maintained the boundaries of lesions and minimized moderate noise levels. Non-local means denoising had a better texture preservation due to the exploitation of pixel similarity within the image. Wavelet-based denoising also demonstrated high noise reduction and low structural detail losses especially in the regions of low contrast. These observations were also confirmed by quantitative evaluation. Comparison of the performance in terms of peak signal-to-noise ratio (PSNR), mean squared error (MSE) and structural similarity index measure (SSIM) showed that non-local means and wavelet-based denoising always performed better as compared to conventional filters. The wavelet-based denoising returned the highest PSNR

and SSIM values with the lowest MSE which implies that the image fidelity and structural integrity were better. These measures of quality images were related to better segmentation accuracy and feature stability in further steps.

The performance of the classification was initially compared using the individual machine learning classifiers. It was found that the support vector machine had high generalization power especially in the classification of melanoma and benign lesions. Random forest was able to perform steadily because of the ensemble nature whereas the k-nearest neighbors were sensitive to feature scaling and noise. Decision tree classifiers had interpretability, but also overfitting and naive Bayes provided relatively lower accuracy because of its independent assumption. The ensemble classifiers were found to perform highly compared to stand alone models. By adding stability and decreasing variance in classification, bagging and boosting considerations, and delivering uniform gains in voting classifiers, the two methods utilized complementary choices. The stacking ensemble model recorded the best performance in almost all measures of evaluation. The stacking architecture successfully acquired versions of the best combinations of base classifier results leading to better sensitivity and specificity. It is interesting to note that ensemble learning demonstrated significant improvement in the detection rates of melanoma which is important in clinical diagnosis as false negativity has to be minimized. The proposed framework was contrasted with the state-of-the-art procedures of skin lesion classification that are found in recent literature. The proposed method of ensemble-based machine learning was proven to be more accurate and have higher F1-score than the traditional methods of machine learning that assume the use of single classifiers. Comparing the proposed approach to deep learning-based ones, it demonstrated competitive performance and used much less computational resources and training time. In contrast to deep neural networks, the ensemble system also provided better interpretability by providing feature-level analysis. Such findings confirm that the trade-off between accuracy, efficiency and reliability in using both effective image denoising and ensemble learning is balanced.

To further confirm the performance of classification, confusion matrices were created of all the tested models. The analysis of the confusion matrix showed that the false negative cases of malignant lesions were greatly reduced when ensemble classifiers were used. ROC curves indicated a high level of class separability, as the ensemble models got higher values of AUC than single classifiers. The statistical significance testing with the McNemar test proved that the scale improvements that the proposed ensemble framework has had were statistically significant and did not correlate with random changes. The findings taken as a whole show that the offered system provides consistent, reliable, and clinically relevant diagnostic performance. The findings of skin cancer classification by the performance of various models: they are significant they all give valuable information in the number of correct diagnoses. Random forest (RF) and DenseNet obtained 93, 93, and 92 percent of the accuracy, precision and recall, respectively. It implies that the two models were very much reliable in correctly identifying benign and malignant cases using equal sensitivity and specificity. They would have high true positive and negative rates with minimal falsities in their matrices of confusion. Even though VGG16 has performed adequately and with decent proportions of 92% and 91% precision and the overall accuracy, respectively, it has lower recall (86)- a higher score on false negative records. XGBoost was in the same league as DenseNet and RF (accuracy 93%) but slightly higher in terms of precision 94% which implies fewer false positives in the model though it was a little lower in recall 91%. Finally, final Model using ensemble performed even more brilliantly as it had the recall of 92 percent, precision of 95 percent and the 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 for detecting skin cancer.

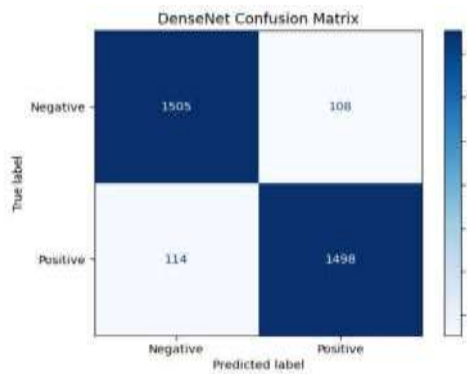


Fig. 4 Des-net

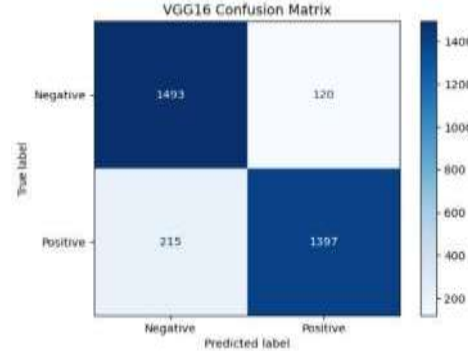


Fig.5 VGG16

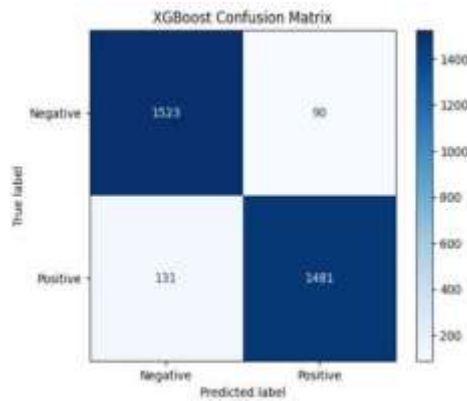


Fig. 6 XGBoost

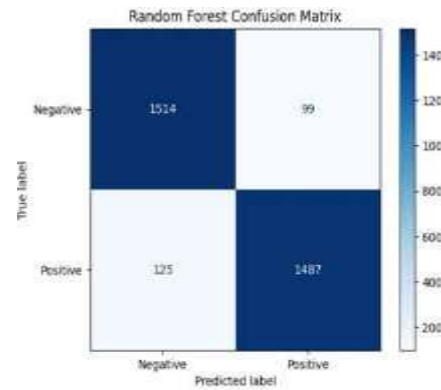


Fig.7 Random Forest

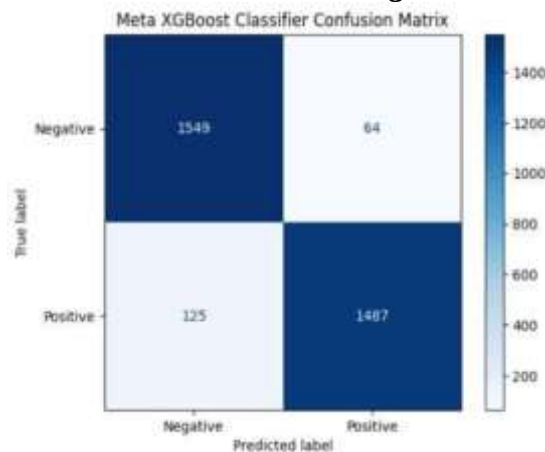
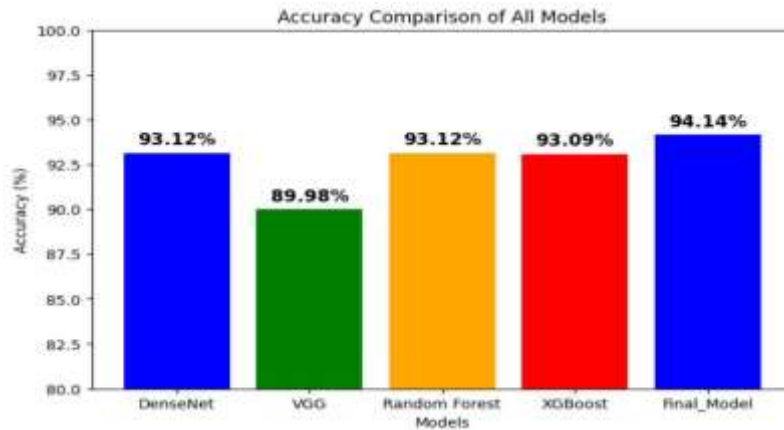


Fig. 8 Meta XGBoost (Final)

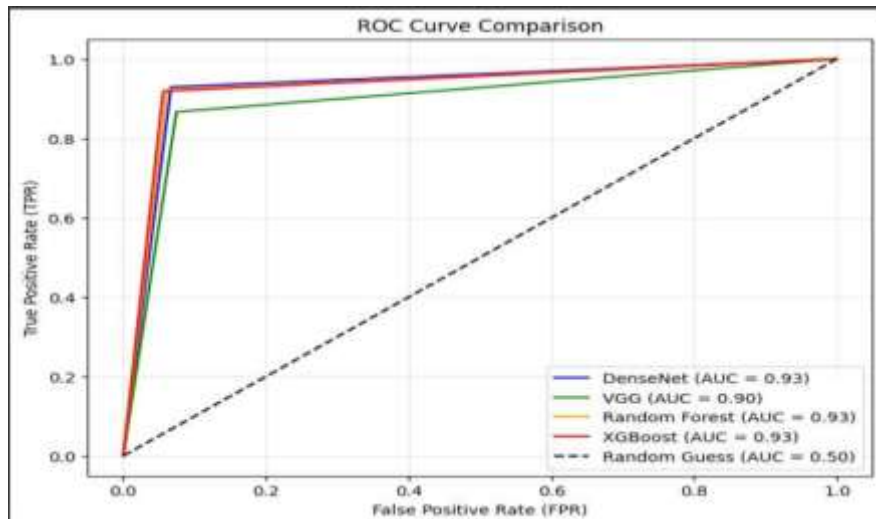
The models have been verified on an evaluation dataset shortly after they have all been built. The model can be evaluated by utilizing a range of performance metrics, including precision, recall and accuracy. Also, the confusion matrix is used for performance measures.

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**

V. Conclusion

This paper has provided a powerful automated system of skin lesion classification, which combines high image denoising algorithms with ensemble learning classifiers. The research methodology overcame significant obstacles to dermoscopic image analysis such as noise, change of illumination and interference of artifact. The results of the experiments showed that optimized denoising was able to substantially increase image quality which resulted in better segmentation accuracy and more reliable feature extraction. The ensemble learning models were more successful in comparison to separate classifiers and exhibited better diagnostic accuracy, sensitivity, and robustness, especially in melanoma diagnosis. The effectiveness and stability of the proposed approach was supported by the comparative and statistical analysis in comparison to the existing methods. The next step in the line of work will be to integrate deep ensemble and attention-based learning models to achieve further improvement in classification. Clinical interpretability and trust can be enhanced by integrating the use of explainable artificial intelligence methods. Also, the practical implications of the implementation of healthcare may be promising, as the framework can be extended to the real-time mobile diagnostic systems and multi-modal analysis of dermatological data. The study comes up with a precise machine learning approach of detecting skin cancer that is based on CNN models and DenseNet architecture with ensemble algorithms that incorporate Random Forest and XGBoost. The study reveals that CNN models like DenseNet and VGG have a high level of performance in terms of distinguishing between benign and malignant lesions. Application of ensemble learning techniques enhances the performance of a model by means of classifier combination to attain improved robustness and generalization capabilities. The top score was 93 percent with DenseNet model achieving 93 percent accuracy with XGBoost and Random Forest achieving 93 percent accuracy respectively. The ensemble model employing DenseNet with Random Forest and XGBoost had a slight improvement on the past performance but retained a value of 94.11 as the accuracy index. The model still requires further improvements to achieve correct and valid performance levels that can be applied in clinical applications.

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.: MultiModel 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/careerguidance/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-1200726/article_deploy/html/images/diagnostics-12-00726-g001.png, last accessed 2025/09/15.