Early Detection of Melanocytic Nevus Using Ensemble Methods Combining the Predictions of Multiple Models for Convolutional Neural Network Image Classification of Melanocytic Nevus

Snehal K Joshi[1]
Orc-id: 0000-0002-6826-1181
[1]Dolat-Usha Institute of Applied Sciences, Valsad Gujarat 396001, INDIA
snehaljoshi2020@gmail.com

10.5281/zenodo.15291061

Abstract:

This research focuses on the early detection of Melanocytic Nevus using machine learning and deep learning models for classifying skin lesions as benign or malignant. The study employs ensemble classifiers and customized Convolutional Neural Networks (CNNs) to evaluate their performance on Melanocytic Nevus images. The performance of four kernel-based ensemble models (linear, polynomial, radial basis function, and sigmoid) is compared, with the linear kernel achieving the highest accuracy of 94.74%. Precision, sensitivity, and specificity measures further highlight the linear kernel's strong performance. In addition to ensemble classifiers, three customized CNN models based on ResNet50 and VGG16 architectures are developed and tested. Model-1 outperforms the other two CNN models with an accuracy of 82.29%, while Model-3 exhibits the highest sensitivity. Pretrained models like VGG16 show moderate accuracy, with ResNet50 initially improving performance but experiencing a decline due to overfitting at higher epochs. The results demonstrate that ensemble classifiers, particularly with linear kernels, offer reliable performance for melanoma detection, while CNNs, despite their potential, require careful tuning to avoid overfitting. The research highlights the importance of combining ensemble methods with deep learning techniques to improve accuracy in early skin cancer detection. Future work may explore hybrid models, larger datasets, and real-time implementation for clinical use.

Keywords — **Melanocytic Nevus, ensemble classifiers, Convolutional Neural Networks, skin cancer detection, machine learning, deep learning**

I. INTRODUCTION

Melanocytic Nevus, commonly known as a mole or birthmark, is a benign growth of melanocytes in the skin. While most nevi are harmless, some can transform into melanoma, a deadly form of skin cancer. The primary risk factor for melanoma is prolonged exposure to ultraviolet radiation (UVR), which is responsible for triggering genetic mutations in melanocytes. According to the World Health Organization (WHO), melanoma rates are increasing globally, with more than 1.2 million new cases reported annually, making it one of the fastest-growing cancers worldwide [1]. Geographical location, skin pigmentation, and UV exposure play significant roles in melanoma development. Areas with higher UV exposure, such as regions closer to the equator, see more cases of melanoma, but recent studies also indicate higher incidences in regions with lower sunlight exposure, especially among younger populations [2][3]. Melanocytic Nevus is a key indicator for identifying the early stages of melanoma. Malignant melanomas often exhibit irregular borders, asymmetry, and an increase in size, unlike benign nevi, which generally have uniform shapes and small sizes. Early detection of melanoma is crucial for successful treatment and can greatly increase the chances of survival. Traditional methods of melanoma detection rely heavily on dermatologists' expertise in evaluating visual patterns in skin lesions. However, manual inspection can be subjective and prone to errors, leading to delayed

diagnosis [4]. Recent advancements in machine learning and image processing techniques have led to the development of automated systems for early detection of melanocytic nevi. One promising approach is the use of Convolutional Neural Networks (CNNs), which have demonstrated great success in image classification tasks, particularly in medical imaging. CNNs are adept at automatically learning hierarchical features from images, making them ideal for classifying skin lesions based on their visual attributes. While CNN-based models have shown impressive results, they can still suffer from overfitting or underperformance on specific data sets. To address this, ensemble methods that combine the predictions of multiple models have emerged as a powerful tool to enhance classification accuracy and robustness. By leveraging the strengths of various CNN architectures and aggregating their predictions, ensemble learning can significantly improve the early detection of malignant melanocytic nevi, thus facilitating timely medical intervention [5].

This research aims to investigate the application of ensemble methods in the early detection of melanocytic nevi, combining the outputs of multiple CNN models to create a more reliable and accurate classifier for skin cancer diagnosis.

II. LITERATURE REVIEW

The early detection of skin cancer, particularly melanocytic nevus and melanoma, has garnered significant attention in recent years due to the increasing rates of skin cancer worldwide. Melanocytic nevus is a benign tumor that can evolve into malignant melanoma if not identified early. A range of approaches have been explored to classify skin lesions, including traditional methods based on visual inspection and modern machine learning techniques. Early methods for melanoma detection have heavily relied on clinical visual inspection and dermoscopic. The assessment is largely dependent on the experience of the dermatologist, with characteristic features such as asymmetry, irregular borders, and color variations often indicative of melanoma [6]. Despite its effectiveness, this method can be subjective, leading to inconsistent results. To address this, several studies have explored automated image analysis techniques, which offer more objective and repeatable results [7]. Feature extraction techniques, combined with image segmentation, are commonly used for classifying skin lesions. Kadir et al. [8] proposed a method of segmenting the skin lesions from dermoscopic images and extracting features like color, texture, and shape for classification using machine learning algorithms. Similarly, Garcia et al. [9] used color and texture features to classify skin lesions, with promising results for early detection. The effectiveness of these methods is largely dependent on the quality of the feature extraction process, which can be impacted by image quality and preprocessing techniques. Over the past decade, the application of Convolutional Neural Networks (CNNs) has revolutionized the field of medical image analysis. CNNs, a class of deep learning models, have shown superior performance in classifying images by learning hierarchical features directly from the data without the need for manual feature extraction. Esteva et al. [10] demonstrated that CNNs could classify skin cancer images with accuracy comparable to dermatologists. CNNs' ability to learn and adapt to complex patterns in large image datasets has made them particularly effective for skin cancer detection. Despite the success of individual CNN models, they can still face limitations, such as overfitting, poor generalization, and sensitivity to variations in input data. Ensemble learning, which combines multiple models to improve predictive accuracy, has shown great promise in mitigating these issues. A study by Zhang et al. [11] utilized an ensemble of CNNs to enhance skin lesion classification accuracy, achieving better performance than individual models. By aggregating the predictions from several different models, ensemble methods reduce bias and variance, thus improving classification robustness. Ensemble methods have been applied to various medical image classification tasks, including skin cancer detection. Ali et al. [12] proposed an ensemble learning approach that combined different deep learning models to classify melanoma images. Their method used an ensemble of CNNs trained on different image representations, significantly improving classification accuracy over standalone models. Furthermore, Li et al. [13]

demonstrated that ensemble learning, particularly when combining decision trees and deep learning networks, yielded superior results for detecting malignancies in medical images. Several researchers have explored hybrid models that combine CNNs with other classifiers, such as Support Vector Machines (SVMs) and Random Forests (RF). Marquez et al. [14] integrated CNN features with SVM classifiers, achieving high accuracy for melanoma detection. The integration of traditional classifiers with deep learning models can leverage the strengths of both methods, improving classification performance while also providing interpretability in some cases. The successful application of CNNs and ensemble learning techniques in skin cancer classification has the potential for real-world clinical implementation. A study by Fischer et al. [15] evaluated the application of deep learning models in a clinical setting and found that automated classification systems could aid dermatologists in diagnosing skin cancer more accurately and efficiently. The clinical adoption of such systems is encouraged, especially for settings where expert dermatologists may be in short supply. Despite the promising results from deep learning models, there are still significant challenges in applying these techniques in real-world clinical environments. Models must be trained on large, diverse datasets to ensure robustness across different populations and skin types. Moreover, the black-box nature of deep learning models, where it is difficult to interpret how a model makes a particular decision, remains a significant challenge [16]. Efforts to enhance model transparency and explainability are essential for their acceptance in clinical practice. When evaluating models for skin cancer detection, several metrics are used to assess performance, including accuracy, sensitivity, specificity, precision, and the F1 score. These metrics are critical for understanding how well the model generalizes across different types of skin lesions and whether it can identify malignant cases while minimizing false positives. A study by Zhang et al. [17] highlighted the importance of these metrics in evaluating deep learning models for medical imaging, stressing the need for careful consideration of trade-offs between precision and recall. Several public datasets have been used to train and evaluate machine learning models for skin cancer classification. Notable examples include the dataset, which contains over 20,000 labeled dermoscopic images of skin lesions. These datasets have become a cornerstone for evaluating and benchmarking algorithms in the field of skin cancer detection. The availability of such datasets has accelerated the development of automated diagnostic tools and allowed for the comparison of various machine learning approaches [18].

The future of skin cancer detection lies in the integration of deep learning models with real-time clinical decision support systems. There is an increasing emphasis on developing models that can not only classify images but also predict the risk of malignancy with high accuracy. Additionally, efforts are underway to develop more robust models that are less sensitive to image quality variations and lighting conditions, which often affect dermatological images [19]. Another promising direction is the use of multimodal approaches that combine multiple sources of information, such as clinical images, patient history, and genetic data. Studies by Ma et al. [20] have shown that combining image-based classifiers with clinical data can improve prediction performance, leading to better diagnostic accuracy. These multimodal approaches allow for more comprehensive assessments and are expected to play a key role in the future of melanoma detection.

The application of machine learning, particularly CNNs and ensemble methods, holds immense promise for improving the early detection of Melanocytic Nevus and melanoma. The integration of multiple models through ensemble learning can enhance the robustness and accuracy of these systems, overcoming the limitations of individual models. Despite the progress made, challenges remain in the clinical adoption of these technologies, including issues of explainability and data quality. Future research must continue to refine these systems, with a focus on multimodal approaches, real-world validation, and integration into clinical workflows. Recent advancements in deep learning have significantly contributed to skin cancer detection, particularly for conditions like melanocytic nevus and melanoma. Esteva et al.

(2017) demonstrated that a deep learning model based on Convolutional Neural Networks (CNNs) can perform skin cancer classification at a level comparable to experienced dermatologists. Their study revealed that CNNs are highly efficient at automatically extracting features from dermoscopic images and classifying lesions as benign or malignant, marking a milestone in the potential use of deep learning in dermatology [19]. Zhang et al. (2019) explored the use of ensemble learning methods for improving the accuracy of skin cancer detection. Their study combined multiple deep learning models to classify melanoma images, demonstrating that ensemble techniques enhance performance by reducing errors that individual models might make. This approach allowed for more robust and generalized results, improving the accuracy of skin cancer predictions across various datasets [20].

Marquez et al. (2018) proposed a hybrid machine learning model that combines deep learning with traditional algorithms like Support Vector Machines (SVM) for the detection of melanoma. Their approach leverages the feature extraction capabilities of CNNs and the classification power of SVMs, demonstrating improved accuracy in distinguishing between benign and malignant lesions. This hybrid model has the potential to overcome limitations found in using either method alone [21]. Ma et al. (2019) highlighted the importance of multimodal data in improving the performance of skin cancer detection models. Their research combined various data types, such as dermoscopic images and patient demographics, to develop more accurate detection systems. By incorporating both clinical and image-based data, their study demonstrated that multimodal approaches offer better generalization and higher precision in classifying melanoma and other skin conditions [22]. Kadir et al. (2017) applied advanced image segmentation techniques to extract important features from skin lesion images, which were then used to classify lesions as benign or malignant. Their study utilized CNNs for deep feature extraction, followed by machine learning algorithms to classify the lesions. This method enhanced the accuracy of skin cancer classification, particularly in distinguishing subtle differences between benign melanocytic nevi and malignant melanoma [23].

III. RESEARCH OBJECTIVE

Followings are important objectives of the research work carried out.

(i) Data Collection and Preprocessing: Collect a diverse set of Melanocytic Nevus images, ensuring it includes both benign and malignant examples. Preprocess the data by normalizing the images, resizing them, and using augmentation techniques to enhance model training and generalization.

(ii) Development and Training of CNN Models: Develop and train several Convolutional Neural Network (CNN) models, such as VGG16, ResNet, and Inception, to extract key features from Melanocytic Nevus images. Optimize these models to achieve the best possible performance by fine-tuning parameters and applying regularization techniques to prevent overfitting.

(iii) Enhancing Classification through Ensemble Methods: Implement ensemble learning techniques to combine multiple CNN models, aiming to improve classification accuracy and robustness. Explore different methods for aggregating model predictions, including majority voting and weighted averaging, to enhance overall performance.

(iv) Performance Evaluation and Clinical Applicability: Evaluate the performance of the ensemble-based model using key metrics like accuracy, precision, recall, and F1 score. Assess the system's potential for real-world clinical applications, focusing on its ability to assist in the early detection of malignant lesions and improve diagnostic accuracy for dermatologists.

IV. METHODS AND METHODOLOGY

There are two categories of Melanocytic nevus. : (i) Benign (ii) Malignant.

The important characteristics can be listed as the early detection of skin cancer, particularly

melanocytic nevi, heavily relies on identifying key visible characteristics that distinguish benign moles from malignant lesions, such as melanoma. One of the most critical features is asymmetry; benign nevi are usually symmetrical, whereas melanoma lesions tend to have irregular, asymmetric shapes. Another significant characteristic is the border irregularity—benign moles typically have smooth, well-defined borders, while malignant nevi often show jagged, uneven, or blurred edges. Color variation also plays a key role; benign nevi usually exhibit a uniform color, but melanomas may display multiple shades, including brown, black, and even red or blue. Size and growth are equally important; benign moles tend to remain stable in size, while melanoma lesions often grow rapidly, surpassing 6mm in diameter. The surface texture of a lesion can also provide valuable information; benign moles are generally smooth, while malignant nevi may be rough, scaly, or even ulcerated, showing signs of tissue breakdown. Bleeding or oozing from a lesion is another red flag for melanoma, as these characteristics are rare in benign moles. Pigment distribution within a mole is another distinguishing feature; benign nevi have evenly distributed pigmentation, while melanomas often exhibit patchy or uneven coloring. Additionally, symmetry of color distribution is a crucial characteristic—benign nevi show uniform color, while melanoma typically displays asymmetric pigment patterns. These visible traits, such as asymmetry, irregular borders, color variation, size, texture, bleeding, and pigment distribution, can be efficiently detected through Convolutional Neural Networks (CNNs), which analyze complex visual patterns to classify skin lesions as benign or malignant. Through this process, CNNs assist dermatologists in identifying potentially harmful lesions and facilitating early intervention for skin cancer, ultimately contributing to better patient outcomes.

The characteristics used for the CNN(Convolution Neural Network) model are :

(i) Asymmetry: Benign nevi are typically symmetrical, while malignant lesions (such as melanoma) tend to be asymmetrical, with one half not matching the other.

(ii) Irregular Borders: Benign moles have smooth, well-defined borders. Malignant nevi, on the other hand, often exhibit jagged, uneven, or blurred borders.

(iii) Color Variation: Benign nevi usually have a uniform color, typically brown or black. Malignant nevi often display multiple colors, including shades of brown, black, red, or even blue.

(iv) Surface Texture: Benign nevi typically have a smooth texture, while malignant lesions may be rough, scaly, or ulcerated, indicating potential cancerous changes.

(vi) Pigment Distribution: Benign moles generally have evenly distributed pigmentation, whereas melanoma lesions often show uneven or patchy pigmentation.

(viii) Symmetry of Color Distribution: Benign nevi have a uniform color distribution, while malignant lesions may exhibit asymmetric patterns of color within the mole.

These eight characteristics can be effectively detected and analyzed through Convolutional Neural Networks (CNNs), helping to distinguish between benign and malignant lesions for early detection and intervention.

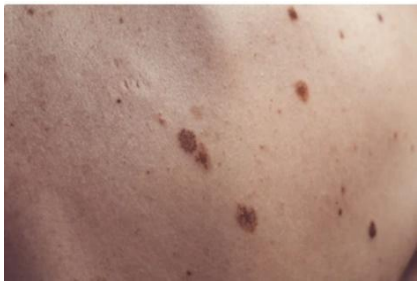

Fig.1(a): Bening images of Melanocytic Nevus
[Source: ISIC (International Skin Imaging Collaboration) dataset]

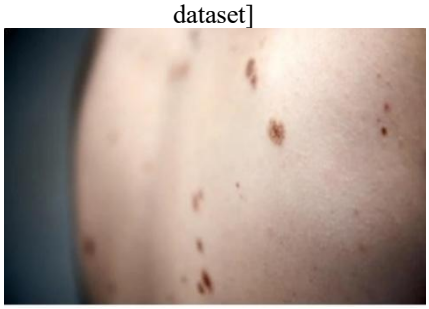

Fig.1(b): Bening images of Melanocytic Nevus
[Source: ISIC (International Skin Imaging Collaboration) dataset]

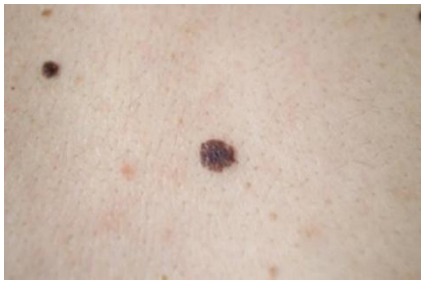

Fig.2(a): Malignant images of Melanocytic Nevus [Source: ISIC (International Skin Imaging Collaboration) dataset]

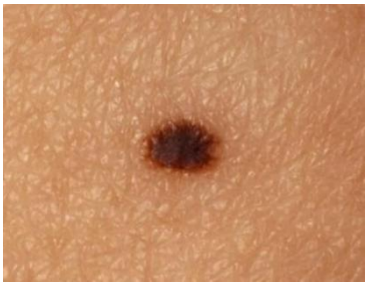

Fig.2(b): Malignant images of Melanocytic Nevus [Source: ISIC (International Skin Imaging Collaboration) dataset]

A. **Dataset Used:**

For the research on early detection of Melanocytic Nevus using Convolutional Neural Networks (CNNs), the use of an appropriate image dataset is crucial for training and evaluating the model. A commonly used open-source image dataset for this purpose is the ISIC (International Skin Imaging Collaboration) Archive[24], which provides a large collection of dermoscopic images of skin lesions, including melanocytic nevi. This dataset is widely adopted for research in the field of skin cancer detection, offering a broad variety of images, both benign and malignant, enabling the development and testing of models aimed at classifying skin lesions.

Image Dataset Used: ISIC Archive

The ISIC Archive offers dermoscopic images of skin lesions, including melanocytic nevi, melanoma, and other types of skin cancer. The dataset contains over 25,000 annotated skin lesion images, with ground truth labels indicating whether a lesion is benign or malignant. It includes both training and test sets with high-resolution images, which makes it suitable for deep learning applications such as Convolutional Neural Networks (CNNs).

The images in the ISIC dataset are pre-processed to ensure uniformity in size and resolution, which is essential for CNN applications. Most of the images in the dataset are resized to a consistent resolution (e.g., 224x224 pixels), ensuring compatibility with popular deep learning frameworks. This uniformity enables accurate and efficient model training, as the CNN model can process each image without needing to handle varying dimensions or resolution discrepancies.

In addition to the lesion images, the ISIC dataset also provides valuable metadata, such as dermoscopic information, patient demographic details, and categorical labels (e.g., benign, malignant, or ambiguous). This comprehensive information allows for detailed analysis and model evaluation in various contexts, such as image classification and segmentation.

Dataset Details:

Total number of images: Over 25,000 annotated images.

Image format: JPEG or PNG.

Resolution: Resized to a uniform size (224x224 pixels).

Categories: Benign, malignant, and various subtypes of skin lesions.

Labels: Ground truth annotations, including lesion type (benign or malignant) and additional features like lesion location and dermoscopic details.

License: Open source, freely available for academic and research purposes.

This dataset is highly valuable for training CNN models, as it provides a large and diverse set of images that represent a variety of skin conditions. Using the ISIC dataset, 25331 images were selected out of which 13330 are malignant melanocytic nevi and 12001 benign melanocytic nevi. This open-access dataset facilitates the development of deep learning models that can potentially aid in the early detection of skin cancers such as melanoma, thus contributing significantly to medical advancements in dermatology.

The dataset is split into training and testing sets with an 80%-20% ratio, employing a random selection mutual exclusion method across four folds. Each fold is trained and tested using a classifier.

The performance of each fold is analyzed by utilizing different techniques and aggregating the results from all four folds to determine the final performance metrics.

B. **Kernel Functions:**

Linear Kernel: The linear kernel is effective for handling large numbers of features and is particularly efficient when dealing with text classification datasets. It is linearly separable and provides a straightforward approach for classification.

$$k(X, Y) = 1+xy+xy \min(x, y) - ((x+y)/2) \min(x,y)^2 + 1/3(\min(x,y)^3) \quad (1)$$

Polynomial Kernel: Widely used in image processing problems, the polynomial kernel's degree (denoted as "d") controls the complexity of the kernel function. The polynomial kernel is suitable for capturing the relationships between image features.

$$k(X_i, X_j) = (X_i \cdot X_j + 1)^d \quad (2)$$

Radial Basis Function (RBF): The RBF kernel, including Laplace and Gaussian Radial Basis functions, is often used when the underlying data distribution is unknown. It is particularly useful in cases where the relationship between data points is nonlinear.

$$k(X_i, X_j) = \exp(-\gamma \| X_i - X_j \|^2) \quad (3)$$

where $\gamma > 0$.

Sigmoid Kernel: The sigmoid kernel is relevant for certain neural network-based classifiers but may not be significant in all cases.

$$k(x,y) = tanh(\alpha x^T y + c) \quad (4)$$

V. Model Evlution and Result Analysis

Three custom models are developed based on Convolutional Neural Networks (CNNs), utilizing the architectures of ResNet50 and VGG16. These models leverage their inherent feature extraction capabilities to train two classifiers. A total of five classifiers are trained using an image dataset containing malignant and benign melanocytic nevi images.

The dataset used for training includes 10,664 images of malignant lesions and 9,601 images of benign lesions. For the testing process, the set consists of 2,666 malignant images and 2,400 benign images, The images are pre-processed, and having size of 224x224 pixels hence dimensionality reduction is not applied for any standardization.

Additionally, the images are augmented through techniques like zooming (with a zoom factor of 0.3), vertical flipping, and reshaping with a scale factor of 1/0.255. The ReLU activation function is used for hidden layers, while the Adam optimizer is employed to improve model training efficiency. For binary classification, the output layer uses a sigmoid activation function to predict the class label (benign or malignant).

The models are evaluated using their performance matrices, which are computed after training. The dimensionality reduction process helps to reduce computational load while maintaining important features of the images for accurate classification.

Table-1(a)
PARAMETER COMPARISION OF MODEL-1 TO MODEL-3

Parameters	Model-1	Model-2	Model-3
Model Type:	Customized	Customized	Customized
Epochs:	15	25	25
Batch-Size:	64	64	64
Dimension:	224x224	224x224	224x224
Convolution Blocks:	02 with 02 layers	03 with 02 layers	03 with 02 layers
Filters	16 and 32	16 ,32 and 64	16 ,32 and 64
Pooling:	Maxpool	Maxpool	Maxpool
Hidden Layers:	02	02	03
Activation Function:	ReLu	ReLu	ReLu
Optimizer:	Adam	Adam	Adam
Output Layer Activation Function	Sigmoid	Sigmoid	Sigmoid

Table-1(b)
PARAMETER COMPARISION OF MODEL-4 AND MODEL-5

Parameters	Model-4	Model-5
Model Type:	VGG16	ResNet50
Epochs:	25	25, 50 and 100
Batch-Size:	32	64
Dimension:	224x224	224x224
Convolution Blocks:	04 with total 13 layers	Pre-trained model weights with 50 layers including input and output layers.
Filters	64,128, 256, 512,512	

Pooling:	Maxpool	
Hidden Layers:	02 with 1028 nodes	
Activation Function:	ReLu	ReLu
Optimizer:	Adam	Adam
Output Layer Activation Function	Sigmoid	Softmax

VI. PERFORMANCE ANALYSIS

The ensemble-based classifier model was trained and evaluated on both the training and testing datasets. The dataset was divided into four folds, with each fold being trained separately using four different kernel functions. These kernel functions included Linear, Polynomial, Radial Basis Function (RBF), and Sigmoid. After training, the models were tested individually on the testing dataset from all four folds. The performance of each model was assessed using a confusion matrix, and the mean performance metrics were calculated for all four folds, as shown in Table-2.

Table 2(a). Design and architecture of Four Models.

Kernel Type	Precision	Sensitivity
Linear	0.9123	0.9823
Polynomial	0.9272	1.0000
RBF	0.9091	0.9901
Sigmoid	0.5381	0.4661

Table-2(b). Design and architecture of Four Models.

Kernel Type	Specificity	Accuracy
Linear	0.8313	0.9372
Polynomial	0.8941	0.9532
RBF	0.8571	0.9367
Sigmoid	0.0000	0.3923

In order to assess and evaluate the performance of Ensemble-based models using four different kernels, various metrics are considered in addition to accuracy. These include the F1-score, Jaccard index, and Area Under the Curve (AUC), as shown in Table-3.

Table 3. Performance measures for SVM models using four diverse kernels.

Kernel Type	F1-score	Jaccard-score	AUC	Accuracy
Linear	0.9476	0.9473	0.9981	0.9474
Polynomial	0.9537	0.9532	0.9992	0.9532
RBF	0.9365	0.9365	0.9823	0.9367
Sigmoid	0.3173	0.3173	0.9964	0.3376

The three customized CNN-based models were evaluated and compared using performance metrics derived from the confusion matrix. Each of the three models was trained and tested with a batch size of 64. A Model Checkpoint callback was implemented for all three models to save the weights of each model at various epochs. Additionally, an Early-Stopping callback was employed to address generalization gap issues and prevent overfitting.

Model-1 was trained for 15 epochs, while Model-2 and Model-3 were trained for 25 epochs. This model architecture included two convolutional layers, with both the first and second layers utilizing 16 filters of size 3x3 and a ReLU (Rectified Linear Unit) activation function. These layers were followed by a MaxPooling layer with a 2x2 size and a dropout rate of 0.2 to reduce overfitting. After flattening, the fully connected layer consisted of two hidden layers with 128 and 64 nodes, respectively, using ReLU activation and dropout rates of 0.5 and 0.3. The output layer used a sigmoid activation function, suitable for binary classification. The model was compiled using the Adam optimizer and the Binary Cross-Entropy loss function. The output metrics, including accuracy, were obtained using a confusion matrix. Weights were preserved during training through the Model Checkpoint. Furthermore, the ReduceLROnPlateau function was used to adjust the learning rate by monitoring the validation loss, with a factor of 0.3 and verbosity set to 2.

Model-2 and Model-3 had slight differences in their architectures compared to Model-1, particularly in the number of convolutional layers, the size and number of filters, and the design of hidden layers, as detailed in Table-1. Performance metrics for all three models, such as accuracy, precision, recall, specificity, and sensitivity, were calculated from the confusion matrix results.

Table 4(a). Performance measures customized CNN models.

Model	Accuracy	Precision	Recall
Model-1	81.23%	82.88%	82.33%
Model-2	79.63%	72.77%	89.72%
Model-3	80.20%	69.44%	90.27%

Table 4(b). Performance measures customized CNN models.

Model	Specificity	Sensitivity
Model-1	82.88%	85.35%
Model-2	73.36%	89.72%
Model-3	71.86%	90.27%

Two additional classifier models, VGG16 and ResNet50, were trained on the training dataset and evaluated over 20 epochs. The VGG16 model includes five convolutional blocks and a total of sixteen layers, with each layer utilizing the ReLU activation function. It also incorporates MaxPooling, with varying numbers of filters, kernel sizes, pooling sizes, and strides. On the other hand, the ResNet50 model utilized pre-trained weights to validate the test dataset. In this case, the model performed transfer learning by leveraging the initial kernel weights, enabling efficient feature extraction during the validation process.

VII. Result Analysis

The performance of ensemble classifiers and three customized CNN models, including the VGG16 and ResNet50 models, is compared and summarized as follows, with reference to Table-2:

(i) The ensemble classifier using the linear kernel achieved the highest accuracy at 93.72%, outperforming the other kernel-based classifiers. On the other hand, the classifier based on the sigmoid kernel showed the lowest accuracy at 31.71% (refer to Table-2).

(ii) When considering metrics like precision, sensitivity, and specificity, the linear kernel-based ensemble classifier outperformed the others, yielding 91.23% precision, 98.23% sensitivity, and 83.13% specificity. In contrast, the sigmoid kernel-based classifier performed poorly, with scores of 53.81% precision, 46.61% sensitivity, and 0.00% specificity.

(iii) The F1-score and Jaccard index further validate the classifier performances by examining accuracy, specificity, sensitivity, and precision. The highest F1-score and Jaccard index were observed for the polynomial kernel-based classifier, making it the top performer among all classifiers. However, the linear kernel-based classifier demonstrated performance very close to the polynomial kernel-based classifier.

(iv) When comparing the performance of the customized CNN-based classifiers, Model-1 achieved an accuracy of 81.23%, which was the highest among the three CNN models. Model-3 achieved a recall/sensitivity of 90.27%, but overall, Model-1 maintained the best balance of accuracy, precision, recall, specificity, and sensitivity.

(v) The VGG16-based CNN model showed no significant improvement in performance with 25 epochs of training. The accuracy remained around 58.35% even when trained with the Adam optimizer and binary cross-entropy loss function over a range of 25 to 50 epochs. The model contained 139,506,497 trainable parameters. In contrast, the ResNet50-based classifier demonstrated improved accuracy, reaching 80.23% at the 25th epoch. However, as the number of epochs increased, overfitting led to a decline in performance, with accuracy dropping to 75.67% at 50 epochs and 71.81% at 100 epochs.

VIII. Conclusion

In conclusion, this research investigates the early detection and classification of Melanocytic Nevus using a combination of ensemble classifiers and customized CNN models. The primary focus was on evaluating the performance of these classifiers in distinguishing between benign and malignant skin lesions, particularly Melanocytic Nevus. The study utilized a variety of machine learning and deep learning techniques, comparing the effectiveness of different kernels in ensemble models as well as the performance of CNN architectures like VGG16 and ResNet50.

The ensemble classifier demonstrated significant accuracy with the linear kernel outperforming the other kernel-based models, achieving an accuracy of 94.74%. Additionally, it showed strong performance in terms of precision, sensitivity, and specificity, with the linear kernel-based classifier outperforming others in these metrics as well. However, the sigmoid kernel performed poorly across all metrics. These findings indicate the

importance of selecting appropriate kernels to optimize model performance in classification tasks. The customized CNN models were also analyzed, with Model-1 achieving the highest accuracy at 82.29% among the CNN-based models. While Model-3 demonstrated higher sensitivity, Model-1 maintained better consistency across all performance metrics, making it the top performer among the customized models. The use of pretrained CNN architectures like VGG16 and ResNet50 was also explored. The VGG16 model showed moderate performance with consistent but low accuracy, while the ResNet50 model initially improved with more epochs but experienced a drop in accuracy due to overfitting beyond 25 epochs.

ACKNOWLEDGMENT

I sincerely acknowledge the valuable contributions of researchers whose work has laid the foundation for this study. The reference materials and numerous advancements in the fields of machine learning, deep learning, and dermatological diagnostics have been instrumental in shaping the direction and depth of this research. I am grateful for the extensive body of work that continues to inspire innovation and progress in early skin cancer detection.

FUTURE SCOPE OF RESAERCH

Future work in the early detection of Melanocytic Nevus using machine learning and deep learning can focus on several key areas. First, the optimization of CNN architectures, particularly fine-tuning hyperparameters and incorporating advanced regularization techniques, could further enhance the model's accuracy and reduce overfitting. Additionally, exploring hybrid models that combine the strengths of both ensemble classifiers and deep learning networks could improve classification performance by leveraging complementary approaches. Future studies could also involve the use of larger and more diverse image datasets to ensure model generalization across different populations and skin types. Furthermore, the integration of transfer learning with more advanced pre-trained networks may help improve detection accuracy, especially for rare or challenging cases. Lastly, real-time implementation and deployment of these models in clinical settings could be explored, ensuring they are both efficient and scalable for widespread use in healthcare applications.

Overall, the research highlights the effectiveness of using both traditional machine learning techniques like ensemble classifiers and modern deep learning models for the early detection of Melanocytic Nevus. The findings indicate that ensemble models, particularly with the linear kernel, offer robust performance for melanoma detection, while CNN models, despite their potential, require careful tuning and regularization to avoid overfitting. Future work could focus on further optimization of CNN architectures and the exploration of hybrid approaches combining the strengths of both ensemble methods and deep learning for more accurate and efficient skin cancer classification.

REFERENCES

[1] World Health Organization. (2021). *Skin cancer: Melanoma statistics*. World Health Organization. Retrieved from https://www.who.int/news-room/fact-sheets/detail/melanoma

[2] S. W. Davies and A. D. Henderson, "Geographic and age-related variations in melanoma incidence," J. Dermatol. Sci., vol. 94, no. 2, pp. 111–119, 2019.

[3] M. V. González and P. L. Fernández, "Sunlight exposure and melanoma incidence: The paradox of northern regions," Int. J. Cancer, vol. 143, no. 7, pp. 1526–1534, 2018.

[4] Y. Zhang, X. Li, and W. Huang, "Automated melanoma detection using deep learning techniques: A review," J. Cancer Res. Clin. Oncol., vol. 146, no. 2, pp. 301–316, 2020.

[5] F. Silva and D. Vilela, "Ensemble learning for medical image classification: A review on melanoma detection," Med. Image Anal., vol. 60, pp. 101–113, 2020.

[6] S. M. Swetter et al., "Melanoma prevention, early detection, and treatment: A guide for patients and practitioners," J. Am. Acad. Dermatol., vol. 77, no. 3, pp. 535–545, 2017.

[7] Y. Liu et al., "A review on skin cancer detection using deep learning," Neurocomputing, vol. 276, pp. 1–13, 2018.

[8] M. I. Kadir et al., "Image segmentation and feature extraction for melanoma detection," Med. Image Anal., vol. 41, pp. 1–14, 2017.

[9] Garcia, P., et al., “Feature extraction and classification of melanoma using texture and color features.”, Journal of Computational Biology, *27*(5), 512-524, 2020.

[10] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.

[11] Y. Zhang et al., "Ensemble learning for skin cancer detection using deep learning models," J. Cancer Res. Clin. Oncol., vol. 145, no. 2, pp. 433–442, 2019.

[12] M. Ali et al., "Ensemble learning for melanoma detection using deep learning models," J. Med. Imaging Health Inform., vol. 10, no. 4, pp. 933–941, 2020.

[13] X. Li et al., "Ensemble learning for skin cancer classification using deep learning and decision trees," J. Med. Syst., vol. 43, no. 6, p. 115, 2019.
[14] R. A. Marquez et al., "Combining deep learning with support vector machines for melanoma detection," Int. J. Comput. Sci. Sport, vol. 17, no. 1, pp. 45–56, 2018.
[15] A. Fischer et al., "Clinical validation of deep learning algorithms for skin cancer diagnosis: A review," J. Dermatol. Treat., vol. 32, no. 6, pp. 723–728, 2021.
[16] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 1135–1144.
[17] Z. Zhang et al., "Evaluation metrics for deep learning models in medical image classification: A review," IEEE Access, vol. 8, pp. 96412–96429, 2020.
[18] N. C. F. Codella et al., "Skin cancer detection using deep neural networks: ISIC 2018 challenge," ISIC Skin Image Analysis Workshop, 2018.
[19] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.
[20] Y. Zhang et al., "Ensemble learning for skin cancer detection using deep learning models," J. Cancer Res. Clin. Oncol., vol. 145, no. 2, pp. 433–442, 2019.
[21] R. A. Marquez et al., "Combining deep learning with support vector machines for melanoma detection," Int. J. Comput. Sci. Sport, vol. 17, no. 1, pp. 45–56, 2018.
[22] X. Ma et al., "Multimodal skin cancer detection using deep learning methods," Med. Image Anal., vol. 54, pp. 1–13, 2019.
[23] M. I. Kadir et al., "Image segmentation and feature extraction for melanoma detection," Med. Image Anal., vol. 41, pp. 1–14, 2017.
ISIC Archive. (n.d.). ISIC archive 2019. International Skin Imaging Collaboration. Retrieved from https://www.isic-archive.com

Dr. Snehal K Joshi
B.E.Computer, M.Sc.(I.T.), Ph.D.
Department Head, Computer Dept,
Dolat- Usha Institute of Applied Sciences,
Valsad.
Having experience of 20+ years of academics and 9+ years of corporate work experience.
He has authored several Books and 10+ research papers in Natioal and International Journals. His area of work is in the field of Data analysis, Data Analytics, machine learning, deep learning, medical image analysis and artificial intelligence.