Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 1, pp. 64–83, DOI 10.12694/scpe.v26i1.3529

DESIGNING AN INTUITIVE HUMAN-MACHINE INTERFACE FOR A SKIN CANCER DIAGNOSTIC SYSTEM: AN ENSEMBLE LEARNING APPROACH

PRASANNA LAKSHMI AKELLA[*]AND R KUMAR[†]

Abstract. In the domain of medical diagnostics, the efficacy of Human-Machine Interfaces (HMIs) plays a pivotal role in harnessing advanced computational models for practical clinical application. This study introduces a refined Ensemble Learning-Based Decision Support System, designed with an emphasis on intuitive HMI for accurate melanocytic and non-melanocytic skin cancer diagnosis. We present "EffiViT," a model that synergizes EfficientNet's robust feature extraction capabilities with the Vision Transformer's attention-based contextual understanding, tailored through an interface that prioritizes ease of use and interpretability for medical professionals. Through extensive evaluation on the ISIC 2019 benchmark dataset, EffiViT demonstrated a classification accuracy of 99.4%, coupled with superior performance in specificity and area under the ROC curve. The system's interface design was iteratively refined based on feedback from dermatologists, focusing on clear visualization of diagnostic information, straightforward navigation, and efficient access to model interpretations. Our findings underscore the importance of integrating user-centered design principles in the development of diagnostic tools, highlighting how a well-conceived HMI can enhance the adoption and effectiveness of AI-based systems in clinical settings. The proposed system stands out not only for its diagnostic accuracy but also for its contribution to the realm of HMI, offering insights into designing interfaces that facilitate better decision-making and ultimately improve patient outcomes in the field of dermatology.

Key words: Skin cancer, Ensemble model, Feature Extraction, Vision Transformer, Data Augmentation, Diagnostic Accuracy, Medical Imaging

1. Introduction and examples.

Skin cancer poses a significant global health concern characterized by abnormal skin cell growth and the frequent emergence of malignant tumors. It primarily manifests in areas exposed to sunlight, often linked to ultraviolet (UV) radiation. Although prevalent in sun-exposed regions, it can also occur in areas with limited sunlight. The three primary types of skin cancer are basal cell carcinoma, squamous cell carcinoma, and melanoma, the latter being the most serious and highly dangerous. The alarming frequency of one skin cancer diagnosis every 57 seconds underscores the need for improved screening techniques and the integration of Computer-Aided Diagnosis (CAD) systems into clinical workflows [1].

In the United States, daily reports indicate over 9,500 new cases of skin cancer [2]. Projections by the American Cancer Society estimate 97,610 new cases of melanoma in 2023, with 7,990 fatalities [3]. Individuals with fair skin, light features, a history of sunburns, or a family history of skin cancer face higher risks. Additional risk factors include compromised immune systems, exposure to chemicals, and radiation therapy. These statistics and risk factors underscore the urgent need for innovative detection and diagnosis techniques, especially for high-risk groups.

Early detection is crucial for minimizing scarring and disfigurement associated with various forms of skin cancer. Therefore, effective management involves techniques for early detection and prevention. Recent advancements in medical image processing, particularly in the identification and classification of skin lesions, have significantly improved diagnostic accuracy [4]. CAD systems address challenges in skin lesion inspection, considering lesion localization, and the presence of hair. These systems aim to assist medical professionals in early detection, automatic identification of malignant lesions, and more efficient treatment.

This study proposes and evaluates a specific Ensemble model to enhance early detection, thereby improving patient outcomes. Recent advances in medical image processing, driven by machine learning and deep learning, have shown remarkable success in diagnosing diseases such as COVID-19 and pneumonia [5, 6]. Both supervised

[*]Department of Electronics and Instrumentation Engineering, National Institute of Technology Nagaland, Dimapur-797103,Nagaland, India, Mail ID: (prasannalakshmi.akella@gmail.com)

[†]Department of Electronics and Instrumentation Engineering, National Institute of Technology Nagaland, Dimapur-797103, Nagaland, India

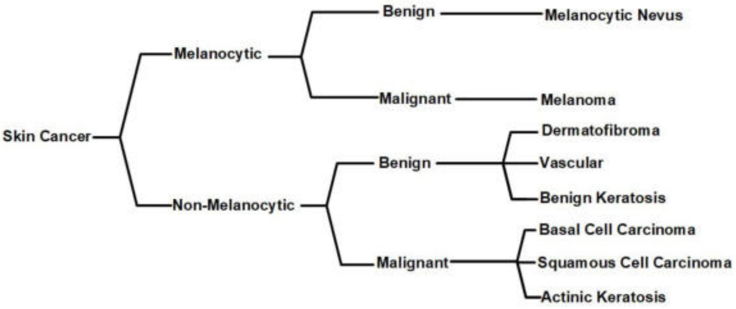

Fig. 1.1: Comprehensive Overview of ISIC 2019 Skin Cancer Types.

and unsupervised methods have been employed in deep learning models for detecting lung and pancreatic cancers [7]. Technologies like AlexNet [8] and VGG16 [9] have demonstrated impressive performance in various applications, including the identification of pulmonary diseases, facial recognition, and unmanned aerial vehicle photography. The various types of skin cancer, each with its unique characteristics, are considered in this study. The proposed Ensemble model, EfficientNetV2 B0-ViT, leverages deep learning technologies to revolutionize the detection and classification of skin cancer, ultimately enhancing patient care and treatment planning. Figure 1.1 illustrates the various skin cancer types examined in the study.

2. Major Contributions and Manuscript Organization. The significant contributions of the EffiViT Ensemble model in the realm of multiclass skin cancer detection are summarized as follows:

- Introduction of the EffiViT Ensemble model, a novel approach combining the strengths of EfficientNet and Vision Transformer (ViT) to advance multiclass skin cancer classification.
- Application of diverse image data augmentation techniques, including rotation, flip, zoom, and noise addition, to enhance the model's classification precision by expanding the dataset.
- Comprehensive evaluation of the model's performance using the ISIC 2019 benchmark dataset, enabling precise comparisons with established methodologies.
- Demonstration of the model's high accuracy in classifying various forms of skin cancer, underscoring its effectiveness for diagnosis and aiding in treatment planning decisions.

The subsequent sections of the paper are organized as follows: Section 3 presents a comprehensive Literature Survey, which critically reviews existing research and advancements in skin cancer classification methodologies. The Materials and Methods, detailed in Section 4, are subdivided into several parts, discussing dataset preparation, preprocessing techniques, data augmentation, and the specifics of the proposed EffiViT Ensemble model that combines EfficientNet and Vision Transformer architectures. Section 5, Results and Discussion, presents a robust evaluation of the model, encompassing experimental setup, performance metrics, confusion matrix analysis, classification reports, ROC-AUC curve analysis, and a comparative analysis with state-of-the-art models. Finally, the manuscript concludes with Section 6, summarizing the study's findings, highlighting the potential of the EfficientNet-ViT Ensemble model in enhancing diagnostic accuracy for skin cancer, and suggesting avenues for future research.

3. Literature Survey. The World Health Organization anticipates that by 2030, cancer will become the predominant cause of death, accounting for an estimated 13.1 million fatalities [10]. Recognizing the global scale of this challenge, research in skin cancer classification has taken on an international dimension, with significant contributions emerging from diverse regions. These studies address the classification across a spectrum of skin types and ethnic groups, underscoring the need for versatile and inclusive diagnostic solutions. Skin cancer is particularly prevalent, arising from abnormal cell proliferation that can swiftly invade and spread throughout the body [11].

A variation of methods for skin cancer classification have been devised and executed in the healthcare field in recent years. For instance, using the HAM10000 dataset, Chowdhury et al. [12] developed a CNN model

Table 3.1: Literature Survey Summary

Study	Method	Dataset Used	Classes	Accuracy
Chowdhury et al. [12]	Custom CNN	HAM10000	7	82.7%
Esteva et al. [13]	CNN	ISIC 2018	7	N/A
Li et al. [14]	VGG16 and ResNet-50	ISIC 2018	7	85%
Nunnari et al. [15]	VGG16 and ResNet-50	ISIC 2019	8	72.2%
Sadeghi et al. [16]	ResNet-50	N/A	4	60.94%
Xie et al. [17]	Deep CNN	ISIC 2017 and PH2	3	90.4%
Yang et al. [18]	ResNet-50	ISIC 2017	2	83%
Zunair et al. [19]	VGG 16	ISIC 2016	2	N/A
Kassem et al. [20]	Google Net	ISIC 2019	8	94.92%
Kasani et al. [21]	Transfer Learning	ISIC 2019	8	92%
Salido et al. [22]	CNN	PH2	2	93%
Shahin et al. [23]	Inception V3 and ResNet-50	ISIC 2018	N/A	89.9%
Sherif et al. [24]	Deep CNN	ISIC 2018	N/A	96.67%
Unver et al. [25]	YOLO and Grab Cut	PH2 and ISBI 2017	N/A	93.39%

that can identify seven kinds of skin diseases. Overall, their technique was 82.7% accurate and 78% precise. In their study, Esteva et al. [13] deployed a Convolutional Neural Network (CNN) to effectively discern seven distinct classes from the ISIC 2018 dataset. Their model achieved an impressive Area Under the Curve (AUC) metric of 94%, indicating its robust performance in accurately classifying the given data. Similarly, Li et al. [14] employed an Ensemble model of ResNet-50 and VGG16 to classify seven skin disease classifications with an accuracy of 85% using the ISIC 2018 dataset.

Nunnari et al.[15] used the ISIC 2019 dataset to classify eight different types of skin. Their rate of accuracy for the explanatory models, VGG16 and ResNet-50, were 72.2% and 76.7%, respectively. Using ResNet-50, Chilana et al. [16] successfully classified 1021 dermoscopy pictures into four skin types with an accuracy of 60.94%. Using a tweaked deep CNN, Xie et al. [17] successfully diagnosed three skin illnesses on the ISIC 2017 and PH2 datasets with an average accuracy of 90.4%. In order to classify two skin illnesses from the ISIC 2017 dataset, Yang et al. [18] employed ResNet-50 and achieved an accuracy of 83%. On the ISIC 2016 dataset, two skin conditions were classified using VGG16 by Zunair et al. [19]. An area under the curve of 81.18% and a sensitivity of 91.76% were obtained.

Kassem et al.[20] on the ISIC 2019 dataset to classify skin lesions into eight categories, proving that image augmentation and transfer learning improve classification accuracy. They achieved accuracy of 94.2%, precision of 73.62% sensitivity of 96.5%, Specificity of 73.62% and F1 Score of 74.04%. After more image enhancement and tweaks to the Google Net's architecture, the accuracy, sensitivity, specificity, precision, and F1 score improved to 94.92%, 79.8%, 97%, 80.36%, and 80.07%, respectively. In order to test several deep learning architectures for melanoma diagnosis, Kasani et al. [21] used image pre-processing to improve image quality and remove noise. To avoid overfitting, they added more data and found that the classification outcomes were considerably improved. 92% precision, 92% recall, and 93% accuracy were all achieved. Classifying skin lesions autonomously was established by Salido et al.[22] A deep CNN improved classification accuracy by 93% and 84% sensitivity by removing noise and artefacts from the image. Shahin et al. [23]used the Inception V3 and ResNet-50 architectures to make a system for classifying skin lesions based on deep neural networks. They trained on the ISIC 2018 dataset and reached validation accuracy rates of up to 89.9 percent. On the ISIC 2018 dataset, the accuracy of the deep CNN used by Sherif et al. [24] to classify and detect melanomas was 96.67%.

For melanoma identification, Unver et al. [25] used the most recent deep learning system. You Only Look Once (YOLO) was utilized for detection, and then Grab Cut was used for cutting out unwanted parts. Using the PH2 and ISBI 2017 datasets, they were able to achieve a precision of 93.39 %.

A summary of literature studies is presented in the table 3.1.

Current study shows a great variety of skin cancer classification methods, emphasizing the importance of machine learning, particularly deep learning. Skin cancer classification accuracy has been improved using

Convolutional Neural Networks (CNNs), VGG16, ResNet-50, and different Ensemble models. The literature reviews observation on methodological variation between research is significant. It's important to highlight that these methodological discrepancies, especially in machine learning model selection and preprocessing, can affect skin cancer classification results. An in-depth discussion about these methodological variations might help explain why some procedures or models yielded better or worse results.

Skin cancer classification is continuously evolving, despite these advances. Current approaches can improve accuracy, precision, and memory, but they still need improvement. Most previous studies have focused on binary classification; however, classification can be expanded towards Multi class classification. In addition to highlighting these disparities, it is important to highlight the obstacles these studies face. Data collection, dataset biases, and overfitting or underfitting during model training can make skin cancer classification model creation and validation difficult. This paper proposes a novel Ensemble model that combines the strengths of EfficientNet and the Vision Transformer architectures to advance the field and overcome these gaps and obstacles. With its promising feature extraction and global dependency capture, this approach aims to improve classification performance. The Ensemble model contributes to skin cancer classification, and exploring other research avenues in the future may further enhance understanding. For instance, improving deep learning models, integrating them with medical diagnostic tools, or exploring new class imbalance methods could lead to beneficial research. In the following sections, the proposed model will be assessed, its efficacy examined, and contrasted with currently used field methods. Despite significant advances in the development of machine learning models for skin lesion classification, the current literature reveals critical limitations. These limitations underscore the need for robust model validation to enhance the accuracy and applicability of these models.

1. *Focus on Binary Classification:* Many studies predominantly focus on binary classification. This approach may not adequately capture the complexities involved in the multiclass categorization of skin lesions, which is necessary to address the diverse nature of skin diseases.
2. *Variations in Methodological Approaches:* There is considerable variability in model selection and data preprocessing across studies. These variations can significantly impact the consistency and accuracy of classification outcomes, leading to potentially unreliable results.
3. *Data Collection and Dataset Biases:* Issues related to data collection and inherent biases in datasets pose significant challenges. These biases affect the generalizability and applicability of the models, making them less effective in real-world scenarios.
4. *Risks of Overfitting or Underfitting:* Many models face risks of overfitting or underfitting during the training phase. This highlights the importance of developing more adaptable and resilient machine learning models that can perform well across various conditions.

4. Materials and Methods.

4.1. Acquiring Dermoscopic Images of Skin Lesions. In this paper, the effectiveness of the proposed method for skin cancer classification is evaluated using the publicly available dataset ISIC 2019. This dataset consists of 25,331 RGB images that offer a comprehensive set of cases for evaluation, spanning eight classes: melanocytic nevus (NV), melanoma (MEL), dermatofibroma (DF), vascular lesion (VASC), benign keratosis (BKL), basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and actinic keratosis (AKIEC). These classes encompass a wide range of skin cancer types, making the dataset an excellent resource for training robust classification models. Table 4.1 presents the class distribution within the ISIC 2019 dataset, including the number of samples for each class. While the ISIC 2019 dataset serves as an excellent resource for this study, its comprehensive array of images and annotations, which span a wide spectrum of skin cancer types, makes it an ideal choice. This dataset not only offers a robust training environment but also ensures that our model is tested against a diverse set of diagnostic scenarios, enhancing its ability to generalize well across different skin cancer classes. However, reliance on this single dataset may limit exposure to variations found in broader clinical settings.

Before training the classification model, the images were standardized to ensure uniform input data quality. This involved resizing all images to 224x224 pixels using bilinear interpolation, normalizing the pixel values, and performing color space conversions when necessary. Additionally, data augmentation techniques, such as random rotations, flips, and zooms, were applied to increase the diversity and variability of the training data. By using augmentation approaches, the problem of overfitting is avoided, improving the model's ability

Table 4.1: ISIC 2019 Dataset Class Distribution

Class & Abbreviation	**Number of Samples**
Melanocytic Nevus (NV)	12,875
Melanoma (MEL)	4,522
Dermatofibroma (DF)	239
Vascular Lesion (VASC)	253
Benign Keratosis (BKL)	2,624
Basal Cell Carcinoma (BCC)	3,323
Squamous Cell Carcinoma (SCC)	628
Actinic Keratosis (AKIEC)	867

to generalize to unseen data. By leveraging the preprocessed and augmented datasets, separate training and testing were conducted on each dataset. The experimental setup involved allocating 80% of the images for training and 20% for validation and testing. This distribution ensures a comprehensive learning process while providing a robust means to evaluate the model's predictive capabilities.

4.2. Data pre-processing. In the framework, thorough data preparation procedures were implemented to assure the best possible state of the multi-class skin cancer dataset for classification accuracy. To commence, each image underwent a resizing process to a uniform resolution of 224 x 224 pixels using bilinear interpolation. This specific size was chosen to balance detail preservation with computational efficiency, making it well-suited for the deep learning models used. The images were resized using bilinear interpolation, which improved their quality and maintained their visual integrity. This step was crucial in preserving the diagnostic features of the skin lesions. To further refine the images, the median filtering technique was utilized to eradicate any noise present in the data.

Median filtering was specifically applied to reduce salt-and-pepper noise, which is common in dermatological images due to variations in lighting and camera quality. Normalizing the pixel values with min-max scaling ensures consistency and comparability across the dataset. Normalization was applied to all channels of the images, adjusting the pixel values to a standard scale that enhances the algorithm's sensitivity to subtle variations in skin lesions. Finally, hair artifacts were removed from the images using a filtering technique known as Blackhat filtering, which effectively obliterated any unwanted hair-like structures. The Blackhat filtering was complemented by a custom algorithm designed to detect and subtract complex hair patterns without affecting underlying lesion details, ensuring that the diagnostic features remain unobscured.

These preprocessing steps optimized the images for subsequent analysis and interpretation by contributing to their enhancement and refinement. Utilizing bilinear interpolation, each image within the dataset was scaled uniformly to a resolution of 224 * 224 pixels. To achieve a balance between the preservation of essential details and the limitation of computational resources, this particular scaling technique was employed. Median filtering was utilized to reduce the presence of objectionable artifacts and noise, thereby improving the image's overall quality. Through the effective reduction of stochastic noise, the application of this technique has resulted in an appreciable improvement in the clarity and coherence of photographs. The goal was to improve the model's ability to extract pertinent data for precise categorization. The pixel values were then normalized using min-max scaling, which effectively rescaled them to suit within the range [0, 1]. This normalization technique has facilitated the accomplishment of consistent training outcomes by standardizing pixel values and fostering convergence across a multitude of features and channels. To reduce the possibility of hair artifacts interfering with accurate classification, a concerted effort was made to eradicate hair. Utilizing Blackhat filtering techniques, the hair filaments present in the epidermis photographs were effectively emphasized and then eliminated.

The primary objective of this phase was to refine the model's focus on the critical patterns associated with skin cancer by eliminating irrelevant data and characteristics. Figure 4.1 illustrates the data preprocessing workflow, showing sample images before and after the application of these techniques. This visual representation underscores the importance of preprocessing in improving image quality for accurate classification. The visual representations are displayed in the left column prior to preprocessing, and in the right column after resizing,

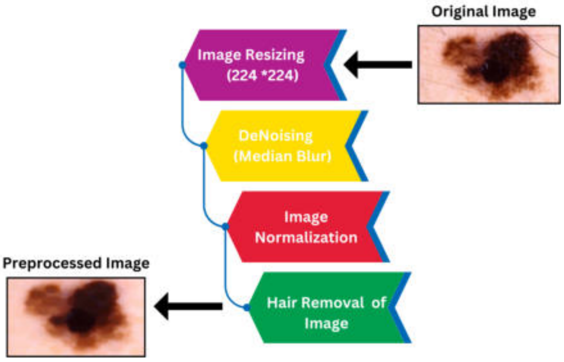

Fig. 4.1: Data Preprocessing Workflow for Skin Cancer Classification. This diagram illustrates the steps involved in preparing data for the analysis and classification of skin cancer, highlighting the crucial preprocessing stages required for effective deep learning model training.

noise reduction, and hair removal techniques have been applied. The provided visual examples demonstrate the effectiveness of employing preprocessing techniques to improve image quality and reduce objectionable artifacts such as hair distortions, among other notable benefits.

The impact of these data preprocessing steps on the subsequent skin cancer classification model's performance was evaluated during the training and evaluation phase. The comprehensive preprocessing workflow aimed to improve model accuracy, robustness, and generalization by minimizing noise, standardizing features, and eliminating hair artifacts. Performing data preprocessing techniques, including resizing, noise removal, normalization, and hair removal, ensured the dataset's quality and suitability for effective multi-class skin cancer classification.

4.3. Data Augmentation. The skin cancer classification model was improved using skin cancer image data augmentation techniques. Data augmentation reduces training data and skin lesion appearance issues. Many augmentation methods were employed to increase dataset diversity and unpredictability. Table 4.2 presents a summary of the original dataset and the augmentation process. Significantly, augmenting the dataset resulted in a substantial increase in the number of images, a pivotal factor contributing to the robustness of the classification model. Random rotations between -10 and 10 degrees and horizontal and vertical flips were applied to create skin cancer photo orientations and mirror image variations. Tiny translations and random zooming simulated scale and viewpoint. These changes improved the model's skin cancer classification. Domain-specific factors determined augmentation methods. Augmentation approaches were customized to highlight relevant changes that match the visual characteristics of different skin lesion types. Highlighting augmentation asymmetry or irregular edges helps train these qualities. Data augmentation categorizes skin tumors empirically. Adding different variants to the dataset has been found to address the problem of insufficient data and increase the model's generalisation skills. Augmented data exposes the model to more skin lesion appearances, helping it acquire robust and discriminative skin cancer features.

Figure 4.2 shows data augmentation-enabled changes in the supplemented dataset. The left column shows original skin lesions, whereas the right column shows augmented ones. These examples show how augmentation tactics create a richer dataset. This part describes the augmentation methods, including translation, zooming, and rotation parameter adjustments. Each augmentation method's explanation and its impact on the model's ability to capture various visual features of skin cancer are explored. Data augmentation was utilized to enhance the skin cancer classification model, aiming to efficiently handle skin lesion appearances and improve classification accuracy.

4.4. Proposed Methodology. The methodology proposed in this study is for the diagnosis of skin cancer entails the application of a multiclass dataset. In order to tackle the difficulties arising from a scarcity of training data and the inherent variability in the visual characteristics of skin lesions, data augmentation techniques are

Table 4.2: Summary of Data Augmentation

Lesion Type	Images Before	Augmentation	Augmented Images
NV	12,875	NO	12,875
MEL	4,522	NO	4,522
DF	239	YES	3,476
VASC	253	YES	4,281
BKL	2,624	NO	2,624
BCC	3,323	NO	3,323
SCC	628	YES	3,423
AK	867	YES	3,476

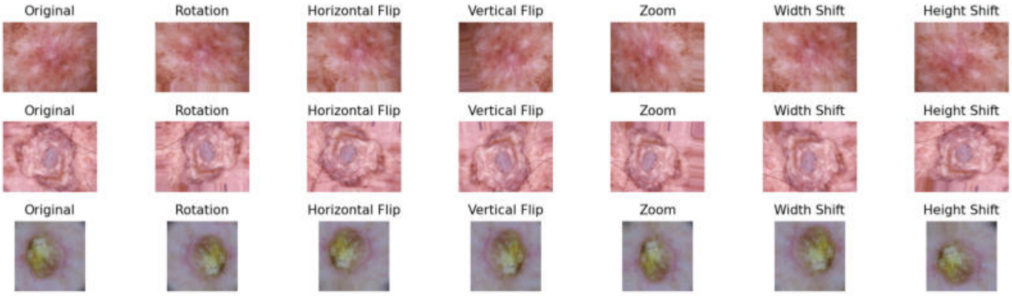

Fig. 4.2: Sample Augmented images

employed. Following the preprocessing stage, two image classification models, namely EfficientNet V2 B0 and ViT-B16, are trained utilizing transfer learning techniques. The augmented dataset was employed to introduce a diverse range of skin lesion presentations, thereby improving the robustness of the classification process. The training procedure includes selecting a loss function with symmetric cross-entropy [27]. Additionally, the Rectified Adam optimizer, which is recognized for its enhanced convergence and efficiency, is selected for optimizing the model.

The Geometric Mean Ensembling technique was utilized to combine the two models into an Ensemble framework, optimizing the strengths of both architectures for superior classification performance. The proposed approach involves assigning suitable weights to the predictions of each model, taking into account their respective performance on the validation set. This strategy aims to capitalize on the unique capabilities of both the EfficientNet and ViT models. The utilization of this Ensemble model allows medical practitioners to leverage computer-aided diagnosis, thereby augmenting the precision and dependability of skin cancer diagnosis. The performance of the model was assessed and measured using a range of metrics, such as Accuracy Score, Precision, Recall, and F1-score, on an independent validation/test dataset. The schematic representation of the entire procedure, encompassing data augmentation, training, and the construction of an Ensemble model, is depicted in Figure 4.3. The proposed methodology effectively encompasses the diverse visual attributes exhibited by skin cancer lesions, thereby augmenting the model's efficacy through the inclusion of a more diverse dataset.

4.5. Models Description of the Proposed Methodology. This section presents a comprehensive description of the models employed in the suggested approach for detecting skin cancer. The Ensemble model synergizes the capabilities of EfficientNet and Vision Transformer (ViT) architectures, aiming to leverage the strengths of each to enhance diagnostic performance.

4.5.1. EfficientNet V2 B0. EfficientNet V2 B0 [26] represents a highly advanced CNN architecture meticulously designed to optimize the process of image classification, particularly in the domain of skin cancer categorization. The model employs Mobile Inverted Bottleneck Convolution (MBConv) and Fused Mobile

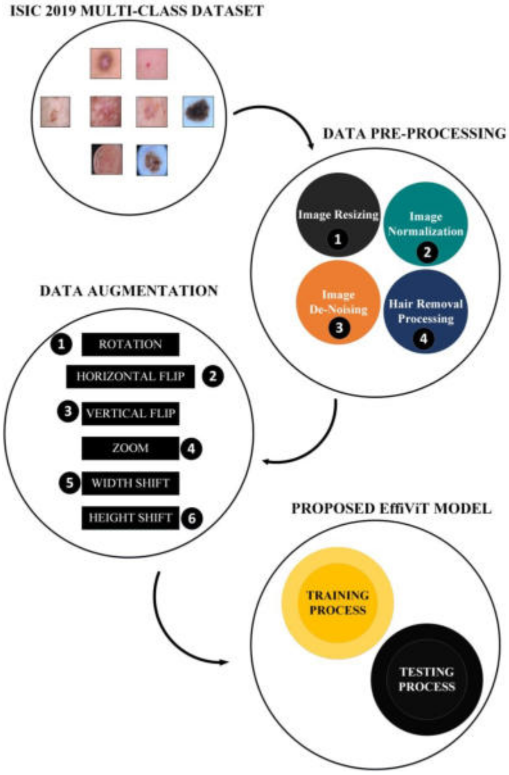

Fig. 4.3: Flow Graph of proposed methodology

Inverted Bottleneck Convolution (Fused-MBConv) blocks, enhancing efficiency and precision. Squeeze-and-Excitation (SE) blocks within these layers enable dynamic importance allocation to different channels, improving discriminative capabilities crucial for identifying skin cancer patterns. For this study, the Efficient Net V2 B0's adaptation involves fine-tuning with the ISIC 2019 dataset, optimizing it for high accuracy in skin cancer classification while maintaining computational efficiency.

Its primary objective is to enhance the efficiency and precision of such applications. EfficientNet V2 is a notable advancement that surpasses its predecessor, EfficientNet V1, by prioritizing the acceleration of training duration and enhancing the efficacy of parameters. The achievement of this objective is facilitated through the integration of compound scaling techniques with training-aware neural architecture search methodologies.

The architecture of EfficientNet V2 B0 incorporates crucial components such as the Mobile Inverted Bottleneck Convolution (MBConv) and Fused Mobile Inverted Bottleneck Convolution (Fused-MBConv) blocks. These elements play a significant role in the overall structure of the model. The MBConv blocks were derived from the MobileNetV2 inverted residual blocks. The architectural design of these blocks incorporates an expansion convolution layer subsequent to a depth-wise separable convolution layer. The Fused-MBConv blocks effectively optimize memory utilization and training duration by integrating the depth-wise and expansion convolutions into a unified standard 3x3 convolution block. These architectural components are indispensable for the extraction and manipulation of features from input images.

EfficientNet V2 B0 incorporates Squeeze-and-Excitation (SE) blocks within both the MBConv and Fused-MBConv layers, enabling the model to dynamically assign importance to different channels and enhance its discriminative capabilities. SE blocks utilize a mechanism called channel-wise relevance weights to dynamically recalibrate feature responses that are specific to each channel. Through the process of recalibration, the model is now able to focus its attention on the most pertinent features and discern crucial patterns linked to skin cancer lesions. EfficientNet V2 B0's design can be effectively elucidated through the use of a visual representation

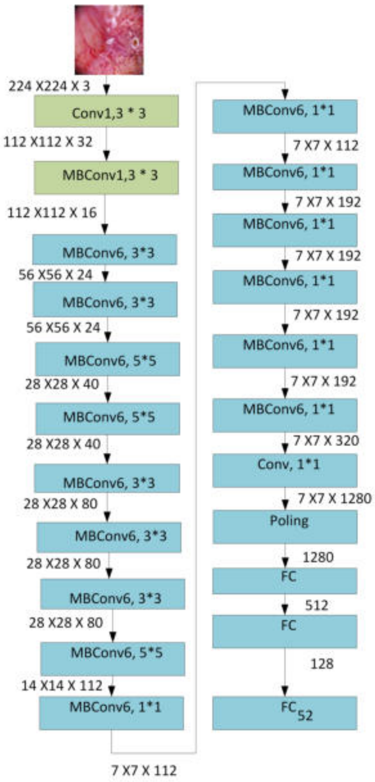

Fig. 4.4: EfficientNet V2 B0 Architecture with SE Blocks for Skin Cancer Classification

shown in Figure 4.4.

The structured visualization of the EfficientNet V2 B0 architecture, as depicted in Figure 5, highlights the sequential layers and data flow within the diagram. The input image is processed by standard convolutional blocks (Conv) in the initial layers. Subsequently, a sequence of MBConv layers is implemented, and these blocks are fundamental to the design of the network. The spatial dimensions and depth of the feature maps adjust in accordance with the data's progression through these layers, as depicted in the diagram. This modification is indicative of the network's compound scaling methodology. The depicted path from unprocessed image input to the ultimate classified output, in which the network completes its assignment of classifying skin cancer lesions, is visually represented in this roadmap. The incorporation of SE blocks into the network architecture, in conjunction with the MBConv layers, augments the model's emphasis on pertinent characteristics, a critical factor in ensuring precise skin cancer detection.

Compound scaling is a pivotal aspect of EfficientNet V2 B0, as it effectively governs the network's depth, width, and resolution. The subsequent information presents the equation for depth scaling.

$$\text{Number of layers in each block} = \text{round}(\alpha \cdot \beta^{\phi}) \tag{4.1}$$

In the equation for depth scaling in the EfficientNet V2 B0 model, each term plays a specific role in determining the architecture of the neural network. The coefficient α sets the baseline number of layers in the network. It acts as a hyperparameter, essentially determining the initial depth of the network. The parameter β is another crucial hyperparameter, typically utilized to control the scaling of the network's width, such as the number of channels in convolutional layers. The factor ϕ is used as a scaling factor, instrumental in adjusting the network's depth. Different values of phi correspond to different versions of the EfficientNet, each version varying in complexity and capacity.

The width scaling formula is given by:

$$\text{Number of channels in each block} = \text{round}(\gamma \cdot \alpha^{\phi}) \tag{4.2}$$

Where γ sets the baseline number of channels (or filters) in the convolutional layers of the network. Together, these equations embody the concept of compound scaling, a distinctive feature of EfficientNet, which ensures a balanced and efficient scaling of the network. This balanced approach optimizes the network's performance while maintaining computational efficiency, a key factor in the success and popularity of EfficientNet models.

4.5.2. Vision Transformer (ViT) B16 model. The ViT model represents a highly advanced methodology for the purpose of image classification tasks. It draws inspiration from the Transformer model, which has gained significant prominence in the field of natural language processing. In this study, ViT B16's application includes a focus on its self-attention mechanisms which are particularly effective in handling the detailed and varied patterns present in skin cancer images. Each encoder in the model uses multi-head attention and a feed-forward layer, enabling it to excel in image classification tasks by recognizing complex interdependencies. ViT B16 has been adapted to analyze skin cancer images, demonstrating its ability to efficiently process and classify large-scale dermatological data.

The ViT model, as opposed to traditional Convolutional Neural Networks (CNNs), leverages self-attention mechanisms to effectively capture extensive dependencies and overarching relationships within images. A multi-head attention layer and a feed-forward layer make up each encoder component. In the multi-head attention layer, attention weights are calculated by comparing how similar different image areas are. The aforementioned procedure can be formally expressed in mathematical notation as follows:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4.3}$$

In the Vision Transformer (ViT) B16 model, the self-attention mechanism is defined by the terms Q (Query), K (Key), and V (Value). The Query represents a specific part of the image, used to determine how much attention other parts of the image should receive. The Key assists in this process by comparing different parts of the image to the Query to calculate attention weights. The Value, representing the actual image content, is then scaled by these weights. This mechanism allows the ViT model to focus on the most relevant features within an image, capturing extensive dependencies and relationships. This approach, diverging from traditional CNNs, underscores the ViT's advanced capabilities in image classification tasks. By applying nonlinear transformations to the attention outputs, the feed-forward layer complements the multi-head attention layer. The formula that represents it is as follows:

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \tag{4.4}$$

In this equation, the input features are denoted by the variable x, and the learnable weight matrices and bias factors are denoted by W1, W2, b1, and b2. The input characteristics are altered by matrix multiplication with W1, then the bias term b1 is added. By introducing non-linearity, the ReLU activation function enables the model to recognize intricate patterns. The generated features then go through another matrix multiplication with W2 before being added to get the feed-forward layer's final output. By applying attention mechanisms and feed-forward layers, Figure 4.5 demonstrates how the ViT model performs significantly well on a range of image classification tasks.

Figure 4.5 represents the ViT B16 model in its initial stages, which involve the creation of image fragments from the input image. In order to preserve spatial information, which is essential for the model to comprehend the arrangement of the image, these regions are subjected to a linear projection in conjunction with positional embedding. In order to extract intricate patterns from the image data, the patches are subsequently processed by the Transformer encoder, which employs layers of multi-head self-attention and normalization. Before being fed into a SoftMax function for the final classification, the output from the encoder is normalized in batches, flattened, and passed through a dense layer.

The system demonstrates exceptional proficiency in capturing and analyzing the interconnections and inter-dependencies present within images. This enables it to effectively discern intricate patterns and accurately

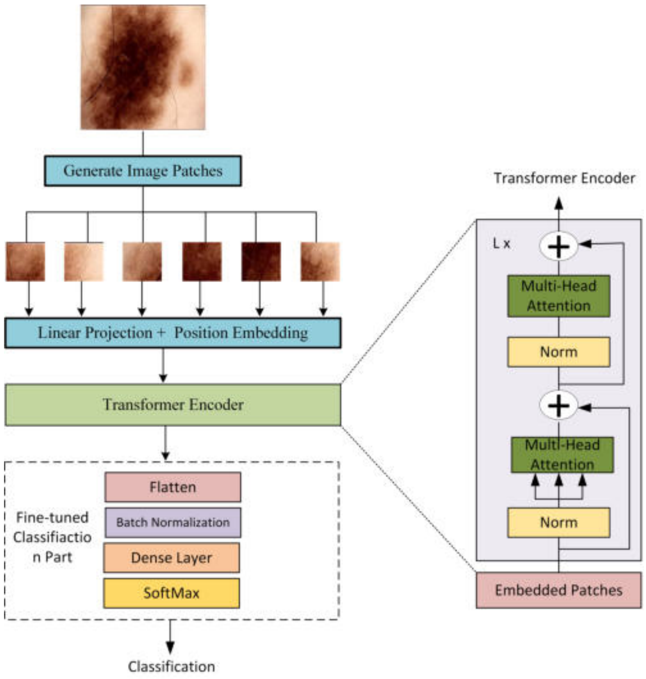

Fig. 4.5: Visualization of ViT Model Performance

classify images based on their inherent characteristics. The ViT model's remarkable scalability and adaptability position it as a preeminent methodology within the realm of computer vision. Its unique characteristics make it especially well-suited for effectively managing vast amounts of information on a large scale.

4.5.3. Ensemble via Averaging: Combining EfficientNet and ViT. In order to classify skin cancer using Ensemble methods, the EfficientNet and ViT Ensemble Model via averaging is a potent strategy to integrate the EfficientNet and ViT architectures. While the ViT excels in capturing comprehensive contextual information from global image regions, EfficientNet has gained significant recognition for its remarkable efficacy and scalability in various computer vision tasks. In this Ensemble model, multiple instances of the ViT and EfficientNet models are independently trained, and diversity is promoted by utilizing various initializations or hyperparameters for each model. During the inference process, the Ensemble generates predictions by employing a technique known as averaging, which involves computing the mean of the individual outcomes produced by each constituent model. This averaging process not only reduces variance among the model predictions but also maximizes the strengths of each architecture, ensuring a balanced approach to feature extraction and context analysis. The ensemble's predictions are further optimized through the utilization of a coefficient weighting approach. The ViT model employs a multiplication operation with a coefficient of 0.7 to scale its output probabilities, while the Efficient Net model utilizes a coefficient of 0.3 for the same purpose. These coefficients were meticulously calibrated based on extensive validation tests that measured the predictive efficacy of each model independently, ensuring that their contributions to the final decision are proportionate to their demonstrated reliability. The coefficients in the ultimate ensemble outcome indicate the proportional significance assigned to the predictions of each individual model.

The Ensemble model effectively consolidates the weighted outcomes to generate a conclusive prediction for the classification of skin cancer. The Ensemble methodology effectively addresses the challenge of reconciling the localized visual data obtained from EfficientNet with the broader global contextual information captured by ViT, primarily due to the implementation of a well-designed weighting strategy. The adjustment of coefficients is contingent upon the problem's inherent characteristics and the performance exhibited by individual models.

The ViT and EfficientNet models have successfully acquired a diverse range of representations through their training process. By leveraging these learned representations, the Ensemble model effectively enhances both the robustness and precision of skin cancer categorization. By implementing a systematic approach that focuses on mitigating the shortcomings of each model, this particular strategy effectively harnesses the unique capabilities and advantages possessed by each individual model.

5. Results and Discussion.

5.1. Experiment Environment. The hardware configuration for the experimental platform utilized in this paper is an Intel Xeon(R) CPU E5-2780 with a 2.80 GHz core frequency and an NVIDIA GeForce RTX 1080 GPU. Using the PyTorch framework [28], the suggested model is implemented in Python 3.7, ensuring a combination of high computational power and state-of-the-art software capabilities for handling deep learning tasks.

5.2. Evaluation Metrics. The evaluation of the classification model incorporated four metrics critical to medical diagnosis: Accuracy, Precision, Recall, and the F1-Score. The selection of these metrics was influenced by their importance in clinical decision-making. In this context, it's crucial not only to achieve high overall accuracy but also to effectively reduce the occurrence of both false negatives and false positives.

The accuracy of the model's predictions is the most fundamental performance metric. It can be defined as the ratio of accurately identified samples that are positive or negative to the total number of samples:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{5.1}$$

Precision is a metric that focuses on the number of true positive predictions relative to the total number of positive predictions. It measures the model's ability to correctly identify positive instances:

$$\text{Precision} = \frac{TP}{(TP + FP)} \tag{5.2}$$

Recall is a metric that measures the ability of a model to correctly identify all positive instances:

$$\text{Recall} = \frac{TP}{(TP + FN)} \tag{5.3}$$

The F1 score is the harmonic mean of Precision and Recall:

$$\text{F1 Score} = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)} \tag{5.4}$$

In the above formulas, TP represents true positive samples, TN for true negative samples, FP for false positive samples, and FN for false negative samples.

5.3. Evaluation of Proposed System. Firstly, the dataset is divided into three different sets: the training, Validation and the test sets. The division is performed in accordance with a predetermined ratio of 80:10:10. During training, a batch size of 32 and a learning rate of 0.0001 were employed to optimize the balance between comprehensive learning and computational efficiency. Both EfficientNet model and ViT model undergoes a training process that involves 50 epochs. The choice of 50 epochs was determined based on preliminary experiments that indicated this was an optimal balance between achieving sufficient model convergence and preventing overfitting, given the complexity of the models and the dataset size.

This approach was optimized to balance thorough learning and computational efficiency. After the completion of the training process, the model parameters were assessed by utilising the test dataset. The EffiViT model, which is an Ensemble model, demonstrated the successful integration of the EfficientNet and ViT models, using their respective strengths. The performance indicators acquired during the training and validation stages offer valuable insights into the model's ability to learn effectively and consistently throughout the learning process. As an example, during epoch 1, the training accuracy of EfficientNet-V2 B0 is recorded as 75.096%,

Table 5.1: Model Performance Summary

Epoch	EfficientNet - V2 B0				ViT- B16				EffiViT			
	Train Acc	Train Loss	Valid Acc	Valid Loss	Train Acc	Train Loss	Valid Acc	Valid Loss	Train Acc	Train Loss	Valid Acc	Valid Loss
1	75.096	1.0863	73.36	0.9125	31.29	1.3972	56.15	1.156	90.115	0.418	93.145	0.158
4	82.65	0.8543	79.46	0.8060	44.67	1.2012	62.83	1.032	92.00	0.350	94.500	0.125
8	84.716	0.7711	80.33	0.7540	57.82	0.9876	68.95	0.908	93.750	0.300	95.750	0.100
12	89.716	0.7149	82.66	0.7312	65.82	0.8542	72.34	0.794	95.250	0.250	96.500	0.085
17	90.145	0.6571	85.007	0.7180	71.18	0.7219	75.12	0.682	96.5500	0.200	97.000	0.070
20	91.635	0.6379	87.75	0.7012	74.56	0.6511	76.98	0.621	97.250	0.180	97.375	0.065
24	93.813	0.5508	90.145	0.6169	78.12	0.5834	79.24	0.562	97.750	0.160	97.625	0.060
28	94.847	0.5067	92.849	0.6010	80.89	0.5234	81.15	0.514	98.000	0.140	97.750	0.055
32	95.936	0.4903	93.183	0.5948	83.24	0.4821	82.72	0.479	98.250	0.120	97.875	0.050
37	96.118	0.4126	94.748	0.5585	85.47	0.4439	84.31	0.447	98.500	0.100	98.000	0.040
40	96.813	0.3893	94.999	0.4028	87.22	0.4123	85.64	0.419	98.750	0.080	98.055	0.030
44	97.145	0.3893	96.318	0.2855	88.75	0.3847	86.78	0.395	99.000	0.060	98.110	0.035
48	97.995	0.2545	97.113	0.2067	90.12	0.3585	87.82	0.372	99.100	0.040	98.165	0.030
50	98.9324	0.1976	97.491	0.1933	91.27	0.335	88.67	0.352	99.20	0.020	98.195	0.015

whereas the validation accuracy stands at 73.36%. In contrast, it can be observed that ViT-B16 demonstrates comparatively lesser accuracies, whereas EffiViT showcases the highest levels of accuracy. As the training advances to 50 epochs, a noticeable enhancement in accuracy and reduction in loss is observed across all models. Particularly, EffiViT consistently exhibits superior performance compared to the individual models.

By following these steps, the proposed method EffiVIT Ensemble model is used for skin cancer classification. This Ensemble model takes advantage of the strengths of both models to enhance classification performance and accuracy. Table 5.1 shows the performance metrics for the Ensemble Model, ViT-B16, and EfficientNet-V2 B0 models during training and validation.

Figure 5.1 illustrates the loss and accuracy curves for both the Vision Transformer model and the EfficientNet model throughout the training and validation phases. The graph on the left illustrates the Training and Validation Loss, wherein both losses exhibit a decreasing trend across the epochs, suggesting a notable enhancement in the model's performance. The EffiViT Ensemble model exhibits the lowest validation loss, indicating superior generalisation capabilities. The graph on the right displays the Training and Validation Accuracy. It is evident that the EffiViT Ensemble model exhibits the best accuracy, suggesting its superior predictive capabilities. The presented graphs highlight the superior learning capacity and effectiveness of the Ensemble methodology compared to the training of individual models.

In addition, the Ensemble model outperforms individual models in terms of inference time efficiency. While the EfficientNet V2 B0 model demonstrated a balance of speed and accuracy, the ViT- B16 excelled in capturing global dependencies, albeit at a slightly higher computational cost. The Ensemble model, through its strategic combination of both architectures, achieves an optimal balance of accuracy and inference time, suitable for real-time clinical applications The EfficientNet -V2 B0 model has an inference rate of 0.0059 seconds per sample and an average inference time of 1.66 seconds per sample, with a standard deviation of 0.93 seconds and 0.0033 seconds, respectively. The ViT-B16 model has an inference rate of 0.0089 seconds per sample and an average inference time of 2.49 seconds per sample, with a standard deviation of 0.08 seconds and 0.0003 seconds, respectively. With an inference rate of 0.0147 seconds per sample, the Ensemble Model, which combines the predictions of the two models, obtains an inference time of 4.15 seconds.

According to these inference time results, the Ensemble Model nevertheless maintains a fair inference rate, making it useful for real-world applications, even though the combining of numerous models makes it slightly more computationally time-consuming. The increased diagnostic performance and rapid inference time of the Ensemble Model demonstrate its potential as a reliable and efficient approach for diagnosing skin cancer.

5.4. Confusion Matrix Analysis. The confusion matrix, as shown in Figures 5.2, 5.3,5.4, offers a comprehensive view of the classification performance of the three models. While overall accuracy assessment is important, the detailed insights provided by the confusion matrix are crucial for understanding the classifica-

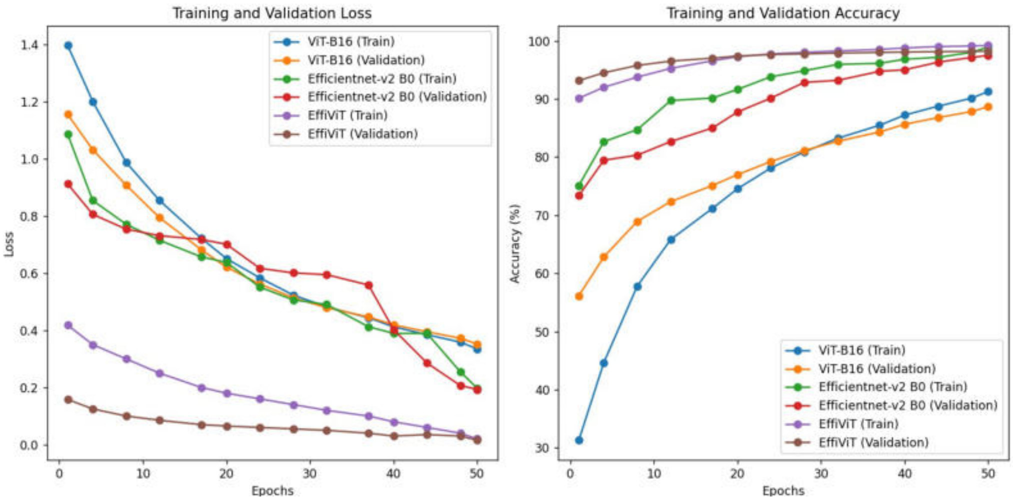

Fig. 5.1: Training and Validation Loss and Accuracy for Models

Table 5.2: Inference Time Efficiency Comparison

Model	**Average Inference Time (s/sample)**	**Standard Deviation**
EfficientNet-V2 B0	0.0059	0.0033
ViT-B16	0.0089	0.0003
Ensemble Model: EffiViT	0.0147	-

tion strengths and weaknesses of each model. The EfficientNet V2 B0 matrix(as seen in Figure 5.2 reveals its precision in classifying NV (Nevus) with 3725 true positives but also indicates a tendency to mistakenly categorize AK (Actinic Keratoses) as NV in 57 instances. Although Melanoma (MEL) and other categories like Dermatofibroma (DF), Vascular lesions (VASC), and Basal cell carcinoma (BCC) are generally well-identified, there are occasional errors, especially in differentiating Actinic keratosis (AK). The confusion matrix of ViT B16(as shown in Figure 5.3 has a continuous pattern characterised by a notable number of true positives for the NV class, along with equivalent performance for the MEL, DF, and VASC classes. Nevertheless, there is a marginal rise in misclassifications, particularly in distinguishing between AK and NV, which is a consistent pattern observed in all models.

The matrix of the EffiViT model(as shown in Figure 5.4 demonstrates notable enhancements, particularly in accurately classifying NV with 3847 instances correctly identified as true positives. Moreover, the model exhibits improved overall performance by reducing the number of misclassifications. It accurately recognizes MEL 1346 times, DF 1038 times, and other classes with great accuracy. It is noteworthy that the accuracy of AK exhibits a little enhancement, so highlighting the Ensemble model's aptitude for distinguishing among increasingly difficult categories. The utilization of a confusion matrix offers a comprehensive evaluation of the efficacy of various models in accurately categorizing distinct types of skin lesions. This analytical tool plays a crucial role in measuring the competency of these models in their classification tasks. The detailed information provided is of great value in formulating therapeutic strategies for the screening of skin cancer. The performance of the Ensemble Model is remarkable, exhibiting an accuracy rate of 99.4%. This surpasses the performance of existing models and signifies a noteworthy progression in the field of skin cancer classification. Furthermore, it has the potential to establish novel benchmarks in terms of diagnostic accuracy. The efficacy of the EffiViT Ensemble model is demonstrated by its capacity to minimize mis-classifications, specifically in distinguishing between AK and NV. This is achieved by leveraging the individual capabilities of the EfficientNet V2 B0 and ViT B16 models, resulting in a powerful combined model.

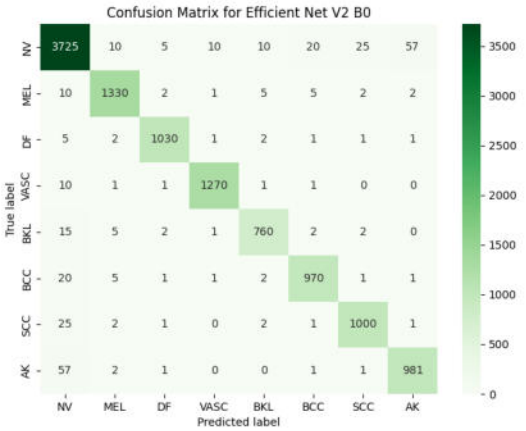

Fig. 5.2: Model Comparison: Confusion Matrix for EfficientNet V2 B0

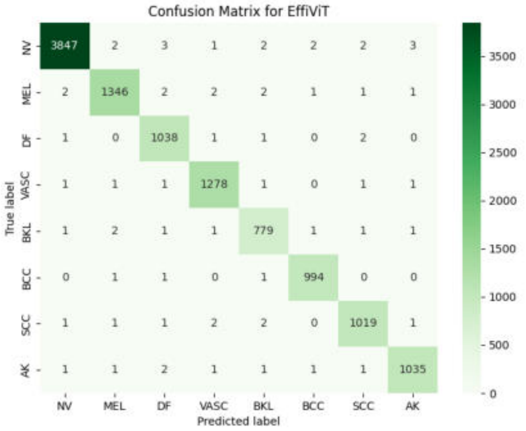

Fig. 5.3: Model Comparison: Confusion Matrix for VitB16

5.5. Classification Report. To offer a comprehensive analysis of performance indicators, including precision, recall, F1-score, and support, Table 5.3 is included, providing a detailed breakdown of these metrics by class. The table presents comprehensive performance metrics for each class across three models, namely EfficientNet-V2 B0, ViT B16, and the EffiViT Ensemble Model. The NV class, which exhibits the highest level of support with a total of 3862 instances, has outstanding precision and recall rates of 99.6% and 99.7% respectively when employing the Ensemble Model. These results indicate the model's remarkable capability in precisely identifying true positives and its reliability in effectively separating the NV class from other classes. Regarding Melanoma (MEL), all models exhibit elevated metrics; however, the Ensemble Model crosses the 99% barrier in precision and recall, indicating a noteworthy decrease in both false positives and false negatives for this crucial category. The performance of Dermatofibroma (DF) and Vascular lesions (VASC) is especially noteworthy, as the Ensemble Model demonstrates somewhat superior recall for DF and precision for VASC. This suggests that the Ensemble Model excels in accurately diagnosing these less common disorders.

The Ensemble Model demonstrates higher precision and memory rates for Benign keratosis-like lesions

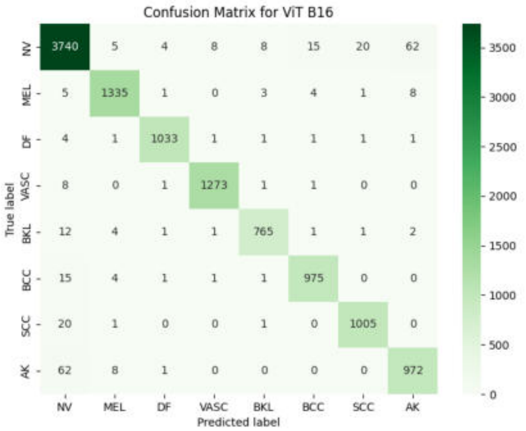

Fig. 5.4: Model Comparison: Confusion Matrix for EffiViT

Table 5.3: Class-Wise Performance Metrics of the Models

Classes	EfficientNet-V2 B0			ViT B16			Ensemble Model: EffiViT			Support
	Precision	Recall	F1-Score	Precision	Recall	F1-Score	Precision	Recall	F1-Score	
NV	96.3	96.4	96.3	96.7	96.8	96.7	99.8	99.6	99.7	3862
MEL	98	98	98	98.3	98.3	98.3	99.4	99.1	99.2	1357
DF	98.7	98.7	98.7	99.1	99	99	98.9	99.5	99.2	1043
VASC	98.9	98.9	98.9	99.1	99.1	99.1	99.3	98.9	99.4	1284
BKL	97.1	96.5	96.8	98	97.2	97.6	98.7	99.6	98.8	787
BCC	96.9	96.9	96.9	97.7	97.7	97.7	99.4	99.6	99.5	1001
SCC	96.8	96.8	96.8	97.7	97.8	97.8	99.2	99.2	99.2	1032
AK	94	94	94	93	93.1	93.1	99.3	99.2	99.2	1043
Micro Avg	96.9	96.9	96.9	97.3	97.3	97.3	99.4	99.4	99.4	-

(BKL), Basal cell carcinoma (BCC), and Squamous cell carcinoma (SCC), with a notable emphasis on BCC. Specifically, the precision rate for BCC reaches 99.6%, while the recall rate reaches 99.5%. The Ensemble Model demonstrates enhanced skill in accurately discerning Actinic Keratoses (AK), a condition that often poses a substantial challenge. It achieves precision and recall rates of 99.3%, indicating its effectiveness in improving the diagnosis of this class.The row labelled 'Micro Avg' presents the mean performance measure across all classes, so providing a comprehensive performance evaluation for each model. The Ensemble Model exhibits a micro-average accuracy, precision, recall, and F1-score of 99.4%, hence illustrating its uniformity across all classes.

The high precision and recall seen across all classes demonstrate the reliability and resilience of the Ensemble Model, which are essential qualities for its potential therapeutic application. The instructive nature of the accuracy rate notwithstanding, it fails to provide a comprehensive understanding of the specific aspects of categorization. To enhance comprehension of classification performance, the confusion matrix, offers a comprehensive analysis of prediction data pertaining to each category. The Ensemble Model demonstrates excellent performance compared to other models, with an impressive accuracy rate of 99.4%. This highlights its exceptional capacity to accurately classify skin cancer in the context of diagnosis.

The comprehensive information provided by the confusion matrix analysis is crucial for comprehending the capabilities of each model in accurately categorizing particular skin lesions, hence directing clinical approaches for the screening of skin cancer. The exceptional performance of the Ensemble Model, with an accuracy rate of 99.4%, represents a notable advancement compared to current models. This achievement has the potential to redefine the benchmarks for accuracy in the categorization of skin cancer.

5.6. ROC-AUC Curve Analysis. In addition to the confusion matrix analysis, the ROC-AUC curve analysis will be a pivotal component of the evaluation. It offers a comprehensive measure of each model's performance across various threshold settings, further confirming their diagnostic reliability and clinical applicability. The ROC curves reported in this study (as shown in Figures 5.5,5.6,5.7) depict the performance evaluation of three distinct machine learning models, namely EfficientNet V2 B0, EffiVit, and ViT B16. The plotted curves illustrate the relationship between the true positive rate (TPR) and the false positive rate (FPR), allowing for an examination of the classifiers' diagnostic capabilities at different threshold levels. The EfficientNet V2 B0 model demonstrates high discriminating capabilities, as seen by its AUC values approaching 1 across all categories. This suggests a remarkable proficiency in distinguishing between positive and negative classes. The EffiVit model exhibits notable performance, as evidenced by some categories achieving an AUC of 1, suggesting the possibility of achieving complete classification accuracy. Finally, the Receiver Operating Characteristic (ROC) curve of the ViT B16 model also exhibits elevated Area Under the Curve (AUC) values, so validating the model's strong precision in tasks related to classification. The persistent positioning of these curves in close proximity to the upper left corner of the graph area signifies a notable degree of precision in forecasting, accompanied by a minimum occurrence of both false positives and false negatives. This highlights the resilience of these models in effectively carrying out their respective predictive functions. Moreover, the models exhibit high AUC values, indicating their suitability for implementation in distinct clinical contexts, such as diagnostic imaging or patient risk assessment, where achieving high levels of sensitivity and specificity is of utmost importance.

The ROC curve is an essential tool for evaluating the trade-off between sensitivity (or TPR) and specificity (1 - FPR) across different thresholds without requiring an arbitrary classification threshold. This makes the ROC curve particularly valuable in medical diagnostic tests where the cost of false negatives varies significantly with the clinical context. This perfect score on the AUC indicates that the model can discriminate perfectly between the positive and negative classes without any overlap. The high AUC values reinforce the potential of these models to act as reliable decision-support tools in medical diagnostics, potentially reducing the cognitive load on healthcare professionals and increasing diagnostic accuracy.

This finding underscores the capacity of these classifiers to serve as decision-support instruments in the field of medical diagnostics, enhancing the proficiency of healthcare professionals. In a comparative analysis, these models demonstrate comparable or superior performance to existing benchmarks in the realm of automated diagnosis, signifying a notable progression in the field of artificial intelligence within the healthcare domain.

5.7. Comparison with State of Art Models. This section provides a comparative analysis of proposed skin lesion classification models using the ISIC 2019 dataset, which is widely recognized as a crucial benchmark in the field of dermatological machine learning research. The Ensemble model, EffiViT, demonstrates superior performance compared to current benchmarks, attaining an overall accuracy rate of 99.4%. The achieved accuracy surpasses the average accuracy of 94.6% reported in the literature for the identical dataset. The individual models, namely EfficientNet-V2 B0 and ViT B16, exhibit strong performance, surpassing the reported average accuracies. The findings of this study highlight the efficacy of integrating sophisticated structures and Ensemble approaches, establishing them as promising instruments for practical implementation in the field of skin lesion diagnosis. Figure 10 visually presents the mentioned findings, emphasizing the superior precision of the models compared to those investigated in the existing literature.

Figure 5.8 illustrates the comparison of accuracy significance of proposed methodological breakthroughs in the classification of skin lesions. Significant progress has been gained in the field of diagnostic precision by utilizing advanced architectures and employing Ensemble techniques, thereby establishing a new standard. The ramifications of these breakthroughs are significant, since they have the potential to greatly enhance diagnostic outcomes and patient treatment within the field of dermatology.

6. Conclusions. In this study, we introduced EffiViT, an Ensemble Learning-Based Decision Support System for skin cancer diagnosis, emphasizing an intuitive Human-Machine Interface (HMI). The system combines the strengths of EfficientNet and Vision Transformer to achieve a classification accuracy of 99.4%, with a user interface tailored for ease of use and interpretability by medical professionals. Our findings underscore the critical role of user-centered HMI design in facilitating the clinical adoption of AI-based diagnostic tools, improving decision-making and patient outcomes in dermatology. Future efforts will focus on enhancing the HMI

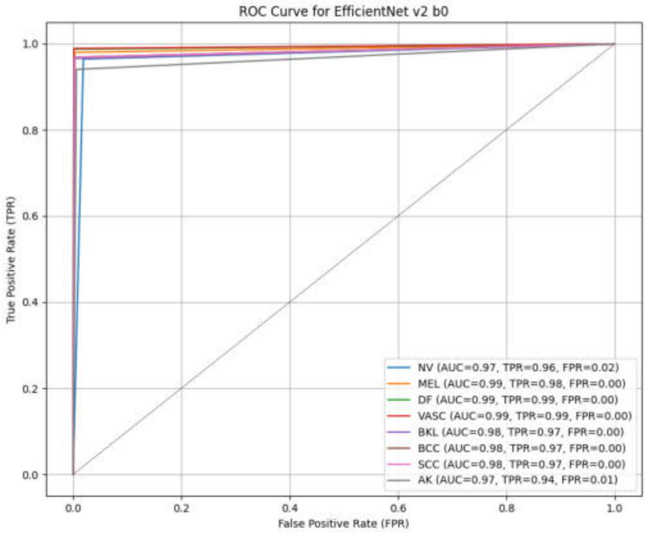

Fig. 5.5: ROC-AUC Curve for EfficientNet V2 B0

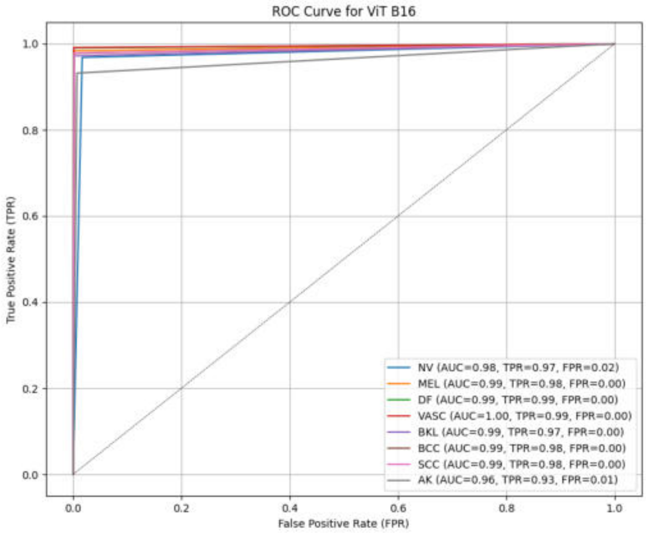

Fig. 5.6: ROC-AUC Curve for VitB16

with more interactive features and extending the system's diagnostic capabilities. By continuing to prioritize the integration of advanced technology with user-friendly interfaces, we aim to further solidify the position of AI as an invaluable asset in healthcare. The success of EffiViT illustrates the transformative potential of combining cutting-edge AI with thoughtful interface design, marking a significant step forward in human-machine collaboration in medical diagnostics.

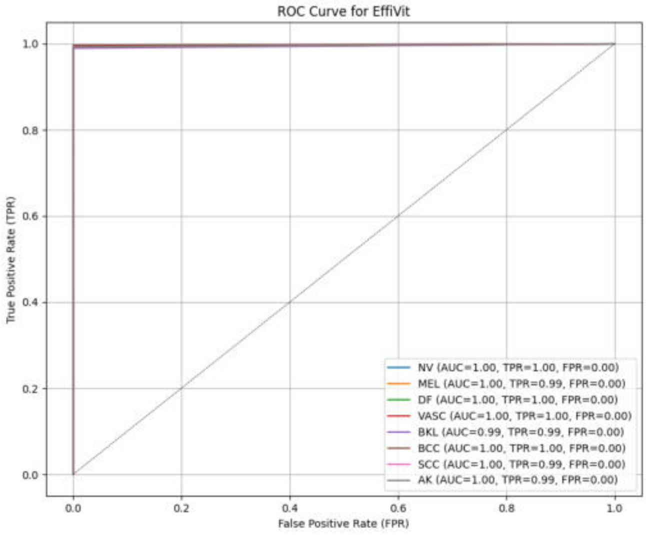

Fig. 5.7: ROC-AUC Curve for EffiViT

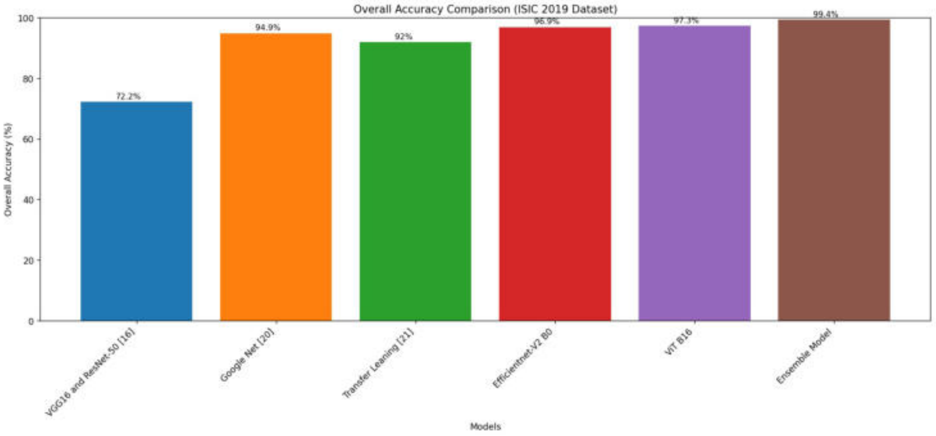

Fig. 5.8: Overall Accuracy Comparison (ISIC 2019 Dataset)

REFERENCES

[1] Andre Esteva, Brett Kuprel, and Sebastian Thrun, *Deep networks for early-stage skin disease and skin cancer classification*, Project Report, Stanford University, 2015.

[2] American Cancer Society, *Cancer Facts & Figures 2022*, American Cancer Society, Atlanta, 2022.

[3] Cancer.Net, *Melanoma: Statistics*, `https://www.cancer.org/cancer/types/melanoma-skin-cancer/about/keystatistics.html`, Accessed 23 June 2023, 2023.

[4] Pedro MM Pereira and others, *Dermoscopic skin lesion image segmentation based on Local Binary Pattern Clustering: Comparative study*, Biomedical Signal Processing and Control, vol. 59, 101924, 2020.

[5] Sammy V. Militante, Nanette V. Dionisio, and Brandon G. Sibbaluca, *Pneumonia detection through adaptive deep learning models of convolutional neural networks*, 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC), IEEE, 2020.

[6] Qing Lyu and others, *Cine cardiac MRI motion artifact reduction using a recurrent neural network*, IEEE transactions on medical imaging, vol. 40, no. 8, pp. 2170-2181, 2021.

[7] Sarfaraz Hussein and others, *Lung and pancreatic tumor characterization in the deep learning era: novel supervised and unsupervised learning approaches*, IEEE transactions on medical imaging, vol. 38, no. 8, pp. 1777-1787, 2019.

[8] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, *Imagenet classification with deep convolutional neural networks*, Communications of the ACM, vol. 60, no. 6, pp. 84-90, 2017.

[9] Karen Simonyan and Andrew Zisserman, *Very deep convolutional networks for large-scale image recognition*, arXiv preprint arXiv:1409.1556, 2014.

[10] M. Manandhar, S. Hawkes, K. Buse, E. Nosrati, and V. Magar, *Gender, health and the 2030 agenda for sustainable development*, Bulletin of the World Health Organization, vol. 96, no. 9, pp. 644, 2018.

[11] R. Erol, *Skin Cancer Malignancy Classification with Transfer Learning*, University of Central Arkansas, Conway, AR, 2018.

[12] T. Chowdhury, A. R. S. Bajwa, T. Chakraborti, J. Rittscher, and U. Pal, *Exploring the correlation between deep learned and clinical features in melanoma detection*, Annual Conference on Medical Image Understanding and Analysis, Springer, Cham, Switzerland, pp. 3-17, 2021.

[13] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, *Dermatologist-level classification of skin cancer with deep neural networks*, Nature, vol. 542, no. 7639, pp. 115-118, 2017.

[14] X. Li, J. Wu, E. Z. Chen, and H. Jiang, *From deep learning towards finding skin lesion biomarkers*, Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2797-2800, 2019.

[15] F. Nunnari, M. A. Kadir, and D. Sonntag, *On the overlap between grad-CAM saliency maps and explainable visual features in skin cancer images*, International Cross-Domain Conference for Machine Learning and Knowledge Extraction, Springer, Cham, Switzerland, pp. 241-253, 2021.

[16] M. Sadeghi, P. K. Chilana, and M. S. Atkins, *How users perceive content-based image retrieval for identifying skin images*, Understanding and Interpreting Machine Learning in Medical Image Computing Applications, Springer, Cham, Switzerland, pp. 141-148, 2018.

[17] Y. Xie, J. Zhang, Y. Xia, and C. Shen, *A mutual bootstrapping model for automated skin lesion segmentation and classification*, IEEE Transactions on Medical Imaging, vol. 39, no. 7, pp. 2482-2493, 2020.

[18] J. Yang, F. Xie, H. Fan, Z. Jiang, and J. Liu, *Classification for dermoscopy images using convolutional neural networks based on region average pooling*, IEEE Access, vol. 6, pp. 65130-65138, 2018.

[19] H. Zunair and A. B. Hamza, *Melanoma detection using adversarial training and deep transfer learning*, Physics in Medicine & Biology, vol. 65, no. 13, Article 135005, 2020, doi: 10.1088/1361-6560/ab86d3.

[20] S. H. Kassani and P. H. Kassani, *A comparative study of deep learning architectures on melanoma detection*, Tissue & Cell, vol. 58, pp. 76-83, 2019.

[21] M. A. Kassem, K. M. Hosny, and M. M. Fouad, *Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning*, IEEE Access, vol. 8, pp. 114822-114832, 2020.

[22] J. A. A. Salido and C. Ruiz, *Using deep learning for melanoma detection in dermoscopy images*, International Journal of Machine Learning and Computing, vol. 8, no. 1, pp. 61-68, 2018.

[23] A. H. Shahin, A. Kamal, and M. A. Elattar, *Deep ensemble learning for skin lesion classification from dermoscopic images*, Proceedings of the 9th Cairo International Biomedical Engineering Conference (CIBEC), pp. 150-153, 2018.

[24] F. Sherif, W. A. Mohamed, and A. Mohra, *Skin lesion analysis toward melanoma detection using deep learning techniques*, International Journal of Electronics and Telecommunications, vol. 65, no. 4, pp. 597-602

[25] H. M. Ünver and E. Ayan, *Skin lesion segmentation in dermoscopic images with combination of Yolo and GrabCut algorithm*, Diagnostics, vol. 9, no. 3, Article 72, 2019.

[26] M. Tan and Q. Le, *EfficientNetV2: Smaller Models and Faster Training*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[27] Y. Wang, X. Ma, Z. Chen, and others, *Symmetric cross entropy for robust learning with noisy labels*, Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, Canada, pp. 322-330, 2019.

[28] A. Paszke, S. Gross, F. Massa, and others, *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, Advances in Neural Information Processing Systems, vol. 32, pp. 8024-8035, 2019.

Edited by: Anil Kumar Budati
Special issue on: Soft Computing and Artificial Intelligence for wire/wireless Human-Machine Interface
Received: Mar 16, 2024
Accepted: May 27, 2024