Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

HDFNet: A Hybrid Deep Fusion Network for Automated Skin Lesion Cancer Detection in Dermoscopic Images

SYED NEHAL HASSAN SHAH[1], IMRAN TAJ[2], SYED MUHAMMAD USMAN.[1],SYED ABDULLAH SHAH[1],ALI SHARIQ IMRAN[3],SHEHZAD KHALID[4]

[1]Department of Creative Technologies, Faculty of Computing and Artificial Intelligence, Air University, Islamabad, 44000, Pakistan email: 222378@students.au.edu.pk, syed.usman@au.edu.pk, 222376@students.au.edu.pk
[2]College of Interdisciplinary Studies, Zayed University, Abu Dhabi P.O. Box 144534, United Arab Emirates email: MuhammadImran.Taj@zu.ac.ae
[3]Department of Computer Science Norwegian University of Science and Technology Gjøvik, Norway email: ali.imran@ntnu.no
[4]Department of Computer Engineering, Bahria University, Islamabad, 44000, Pakistan (email: shehzad@bahria.edu.pk

Corresponding author: Ali Shariq Imran (e-mail: ali.imran@ntnu.no).

ABSTRACT Skin cancer, characterized by the abnormal growth of skin cells, remains a significant global health challenge. The primary types of skin cancer are basal cell carcinoma (BCC) and melanoma (MEL). BCC, although the most common, typically grows slowly and seldom metastasizes. In contrast, melanoma, though less common, is highly aggressive and can spread to other organs if not detected early. This condition is particularly prevalent in regions like Australia, where high sun exposure contributes to the highest incidence rates worldwide. Early detection and accurate diagnosis are crucial for effective treatment. Advancements in Artificial Intelligence (AI) have opened new avenues for improving skin cancer classification. Current AI models, however, often face challenges related to interpretability, generalizability across diverse skin types, and the integration of clinical context. Our study introduces the Hybrid Net Dense Fusion Model, a novel approach that leverages multi-layered deep learning techniques to enhance the classification of dermal lesions. By incorporating attention mechanisms and fusion strategies, our model aims to improve diagnostic accuracy and provide detailed analysis for early detection and treatment planning. The performance of our model, achieving an accuracy of 92% and an AUC-ROC curve of 92.15%, demonstrates significant improvements in diagnostic accuracy. The deployment of this model within healthcare systems promises substantial advancements, offering valuable support to dermatologists, reducing diagnostic times, and enabling earlier interventions. Our model's adaptability and enhanced capabilities highlight the potential of modern AI in medical imaging, ultimately aiming to enhance patient care and outcomes. Our model evaluates multiple feature extraction methods, combining the strengths of ResNet-50, VGG-16, and Vision Transformer (ViT) architectures. By utilizing a dense fusion approach, we integrate features extracted from these models to create a comprehensive representation of dermoscopic images. This method leverages the specific advantages of each model—ResNet-50's robust feature extraction, VGG-16's simplicity and effectiveness, and ViT's ability to capture global context through self-attention mechanisms. Incorporating attention mechanisms further refines the model's focus on relevant areas within the images, enhancing interpretability and precision. The evaluation metrics include precision, AUC-ROC, and F1-score, providing a holistic view of the model's performance. The achieved AUC-ROC score of 92.15% underscores the model's capability to distinguish between cancerous and non-cancerous lesions effectively. The integration of our model into clinical practice can lead to significant improvements in diagnostic accuracy and efficiency. By automating parts of the detection process, it offers valuable assistance to dermatologists, reducing diagnostic times and facilitating early intervention. The model's adaptability across various skin types and its advanced capabilities exemplifies the transformative potential of AI in medical imaging, aiming to improve patient care and outcomes.

INDEX TERMS Skin Cancer, Basal Cell Carcinoma, Melanoma, Artificial Intelligence, Deep Learning, Dermoscopic Images, Attention Mechanisms, Model Evaluation.

I. INTRODUCTION

SKIN cancer is one of those evasive diseases with a variety of identifiable patterns which due to the variations in the skin type itself lead to many cases of false positives or false negatives. There are a number of other factors that may contribute to this evasive behavior such as image noise, light intensity, angle, etc. Through the aid of latest innovations in Artificial Intelligence (AI) and deep learning (DL) models, it is now possible to revolutionize the field of medical imaging through the deep impact of machine learning on the diagnostic abilities of dermatology [1]. Melanoma is one category of skin cancer that is of serious concern due to its potential to progress into fatal levels. This type of cancer is a result of excessive ultraviolet (UV) radiation exposure which leads to melanocytes developing cellular mutations and transforming into malignant tumors. The aid of AI in the detection of such a critical variant of skin cancer will provide invaluable assistance to dermatologists worldwide. However it will not be a wise choice to completely rely on automated systems because while the early detection of skin cancer can certainly lead to timely impediments and treatments that may obstruct its further aggressive development, the patient does also requires an empathic connection with sympathetic human beings to anchor their psyche through this excruciating, painful time. No machine can provide this emotional connection at the moment and therefore it is necessitated that mechanical precision be used in conjunction with human interaction and compassion for the perfect treatment in a synergistic manner.

Before the advent of AI and DL, skin cancer detection typically relied on the observation skills and the expertise of the dermatologists, therefore the accuracy of this visual analysis could be compromised due to any number of factors like brightness, angle of observation, experience of the observer, causing potential oversights that could lead to problematic diagnoses [2] [3]. Through the integration of machine learning, we can utilize a wide variety of architectures like CNN, DNN, RNN and LSTM for lesion detection and identification [3] [4], along with the usage of models such as ResNet, Inception, Xception and VGG16 for enhanced image detection and classification [5]. Among these AI architectures, enhanced InceptionV3 and VGG16 are known to excel in lesion classification when combining human expertise with machine accuracy [3]. Dermoscopy is an important tool for acquiring images that will allow for an accurate diagnosis as dermoscopic analysis is the preliminary step in the assessment of any suspicions regarding the development of cancerous regions [6] though it has been noticed that the detection rate for melanoma remains between 75%-84% [7] despite the visual inspection being aided by dermoscopy which provides a magnified and illuminated view of skin so that any irregularities can be easily discovered. In contrast, traditional DL models significantly increased the detection rate through a more accurate assessment of skin condition. [8]. The recent developments of new models that can discern minute features without falling into any bias have been fruitful for the field of medical imaging due to lesions being difficult to classify [9] however the hurdles that impeded visual inspection still persist to some extent such as skin type variations, presentation method of the affected region and the inherent subjectivity that ensure that a thorough dermatological analysis is requisite in addition to the latest, most suitable architectures [10]. The several architectural families that have been brought to prominence in recent studies such as ResNet, VGG, Inception and Xception have shown exceptional adaptation capabilities for dermatological data through the Few-Shot Learning techniques. Our goal in this study is to assess the compatibility of the VGG series through studying the outcomes of VGG-11, VGG-13, VGG-16, and VGG-19 [11] [12] [13] [3] so as to identify the most accurate skin cancer classification model.

In this study we are have enlisted the set objective that we will contribute from our study which are as:

- To enhance model generalization, methods will be evaluated to improve model. generalization beyond training datasets and enable reliable performance on unseen data.
- To test the developed models on different datasets, performance will be evaluated as in practical scenarios.
- To explore ensemble and transformer learning techniques, efforts will be made to further improve the model's capability to classify skin cancer accurately."

II. LITERATURE REVIEW

We have compared the existing literature on skin cancer detection using dermoscopic images. A typical method involves preprocessing, feature extraction and classification. In preprocessing, Mukadam [9] employed image resizing, color preservation, sharpening filters along with augmentation techniques like rotation, zooming, height and width shift, and rescaling while applying ESRGAN for enhancement while Al-Rasheed [12] used resizing along with rotation, translation and flipping augmentation techniques. Resizing, black hat filtering, masking filtering, noise removal and augmentation methods of rotation, flipping and blurring were applied by Thanka [14] on their data while Kaya [2] did not elaborate on their preprocessing techniques. Jaisakthi [11] used image standardization and resizing while Yang [15] used resizing, normalization, duplication removal, patch extraction, position embedding and augmented through techniques like brightness/contrast alteration, flipping zooming, rotation and shift. Zhao [16] used median frequency balancing, resizing normalization and augmentation while Kahia [17] used class balancing, resizing and augmentation methods of flipping, shifting, rotating, transformation, and zooming. Ali [18] used resolution scaling, hair removal methods like black hat transform, masking and fast marching method, and augmentation techniques like rotation, zooming and horizontal and vertical flipping. Gouda [5] utilized oversampling, resizing, custom contrast method, and augmentation techniques like rotation, reflection, shifting and brightness adjustment, along with ESRGAN for enhancement and noise reduction. Bassel [3] used resizing while Nakai [10] used lesion segmentation, normalization, resizing, and data augmentation.

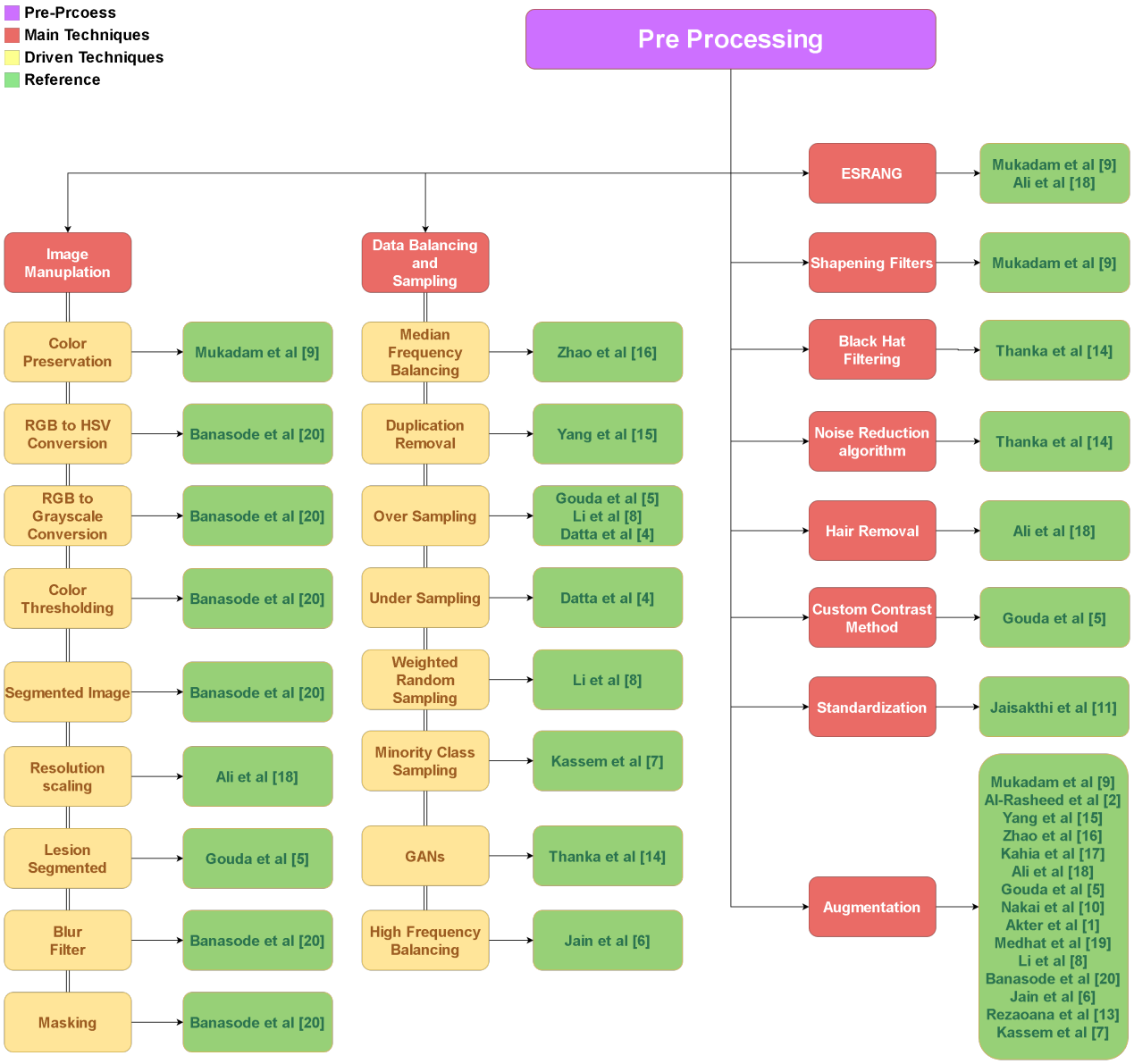


FIGURE 1: Illustrates the preprocessing techniques applied in the previous studies

Akter [1] used resizing, normalization and augmentation while Medhat [19] used resizing and augmentation. Li [8] employed augmentation, oversampling and weighted random sampling whereas Banasode [20] applied image blurring, color thresholding, image masking, segmenting and transformation along with RGB to HSV and RGB to Grayscale conversions for their data. Datta [4] used oversampling, undersampling, and normalization while Jain [6] utilized high frequency sampling, normalization and augmentation methods of rotation, shifting and zooming. Rezaoana [13] employed normalization, resizing, and augmentation through rotation, horizontal flip, zoom and shear. Finally Kassem [7] employed minority class sampling, augmentation, and bootstrapped multiclass SVM aggregation.

In feature extraction, Al-Rasheed [12] utilized transfer learning models ResNet50, ResNet101 and VGG16 pre-trained on ImageNet for feature extraction while Thanka [14] utilized the transfer learning VGG16 model. Jaisakthi [11] used ResNet50, DenseNet121, Inception ResNetV2 and EfficientNet variants 0-7 for feature extraction purposes whereas Yang [15] relied upon patch-based feature representation. Bassel [3] employed ResNet50, VGG16 and Xception (Main) for this stage while Akter [1] used ResNet50, VGG16, Inception, Xception and MobileNet. Banasode [20] used KNN mining/clustering algorithm whereas Kassem [7] employed GoogleNet for feature extraction purposes. These are all the researches that have utilized feature extractors and so we will

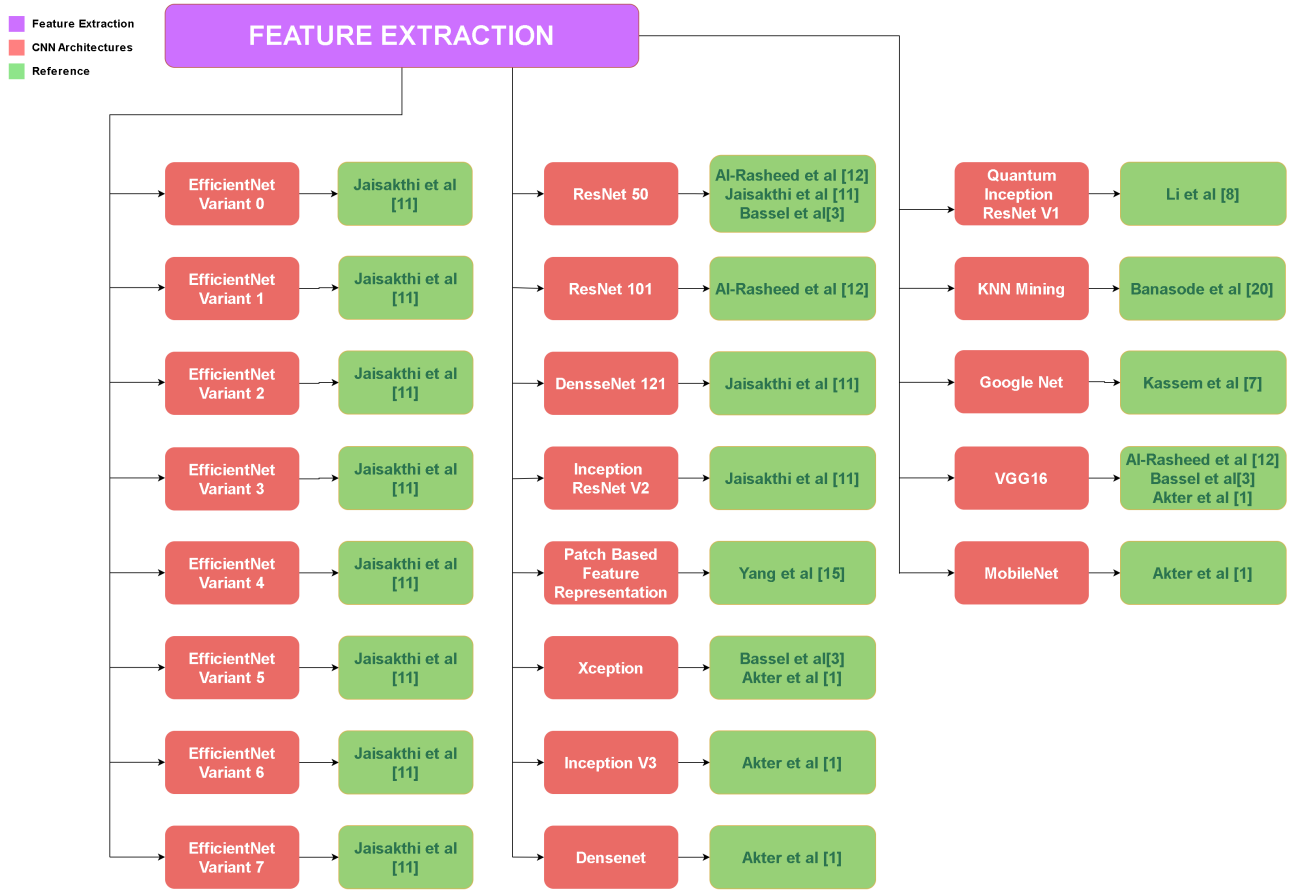


FIGURE 2: Illustrates the Feature Extraction techniques applied in the previous studies

move to discuss the variety in the selection of classifiers that were employed by the previous researches.

Mukadam [9] utilized a custom CNN for seven-category classification whereas Al-Rasheed [12] employed VGG16, ResNet50 and ResNet 101 pre-trained on ImageNet along with CGANs and the Ensemble algorithm for the purpose of generating more data and averaging outcomes respectively. Thanka [14] employed a hybrid strategy involving data augmentation, and Conditional Generative Adversarial Networks (CGANs) with VGG16 and XGBoost while Kaya [2] compared the performance of VGGNet11, VGGNet13, VGGNet16, and VGGNet19 and discovered VGG11 to be the best performer. Jaisakthi [11] used the EfficientNetB6 pre-trained on ImageNet for a binary classification (benign or malignant lesions) while Yang [15] used a custom Vision Transformer for a seven-category classification. Zhao [16] employed a number of CNNs including ResNet18, VGGNet, ViT and Deep ViT; they found ResNet18 to be the best performing model among these. Kahia [17] employed VGG16 and InceptionV3 with transfer learning for two-category and three-category classification purposes whereas Ali [18] applied EfficientNet B0-B6 family of models for a comparative seven-category classification and found EfficientNetB4 to be the best performing model. Gouda [5] selected ResNet50, InceptionV3 and InceptionResNet for a binary classification experiment and concluded with InceptionV3 being the best performer while Bassel [3] used DNN, SVM, Random Forest (RF), Regression, Neural Netowrk (NN), k-nearest neighbours (KNN), AdaBoost, Decision Tree, GaussianN combined with Xception feature extraction using stacking CV algorithm for a binary classification task.

Nakai [10] used an enhanced deep bottleneck transformer (EnDBoT) model, ResNet50 and DenseNet201 for multi-class categorization task. Akter [1] used five distinct stacking models for a seven-category classification whereas Medhat [19] employed MobileNetV2 and AlexNet both with transfer learning and augmentation for a multi-class categorization experiment. Li [8] used quantum computing with InceptionResNetV1 for melanoma classification through SVM classifier while Banasode [20] used the SVM for a binary classification. Datta [4] employed soft-attention layers in ResNet34, ResNet50, VGG16, VGG19, Inception-ResNetV2 and DenseNet201 for multi-category classification of melanoma while Jain [6] employed transfer learning models including VGG19, InceptionV3, InceptionResNetV2, ResNet50, Xception and MobileNet for melanoma classifi-

FIGURE 3: Illustrates the the main classification models applied in the previous studies

cation in multiple categories. Rezaoana [13] utilized parallel CNN model for a nine-category classification whereas Kassem [7] used transfer learning GoogleNet and replaced Softmax and classification output layers with SVM for lesion classification in multiple categories. Several gaps have been identified from the existing literature which are listed as follows:

- Many studies rely on biased datasets, lacking broad strategies to address these imbalances and resulting biases.
- Models often demonstrate strong performance on familiar datasets, yet they frequently struggle when faced with new data, indicating a tendency towards overfitting.
- Most models are tested using the same kind of datasets, bypassing the need for evaluation on varied, real-life skin cancer examples.
- The potential of transformers and ensemble models in skin cancer detection hasn't been deeply investigated which leaving gaps in understanding their full benefits.

III. OVERVIEW OF PROPOSED METHODOLOGY

The brief discussion regarding the prior researches should serve to provide a glimpse of the vast array of possibilities with machine learning architectures and techniques. From this point onwards, we will describe our methodology regarding the preprocessing steps that have been taken for dataset preparation along with elaboration upon our choice of classifiers and accompanying methods, all selected for inclusion with the goal of achieving better performance indicators than before while tackling the difficulties of class imbalance so as to present an effective, robust and generalizing model with real-world applications.

IV. PREPROCESSING

Concerning the choices of operations made upon the acquired data before any experimentation through AI architectures, we begin our preprocessing by firstly compiling all of the sourced data into a single folder so that accessing the input data is intuitively easy. The different datasets that we have sourced for our usage include ISIC-2019, ISIC-2020, PAD-UFES and DermQuestDermIS. Before all the images have been processed and passed to load into model we first standardize the to be applied data in to a fixed size of 150x150x3. Then the input data was loaded into TensorFlow while adjusting the hyper parameters to be 150x150 for image resolution, 10x10 for patch size, class names (BKL, NV, MEL, BCC), dropout rate of 0.1, and a batch size of 32. Manual filtering was applied on the dataset to employ human judgment to ensure that the images collected are suited for utilization as input for classifier tasks. Standardization was used to reduce the different resolutions and formats of the acquired image data to a single resolution (150x150) and format which will

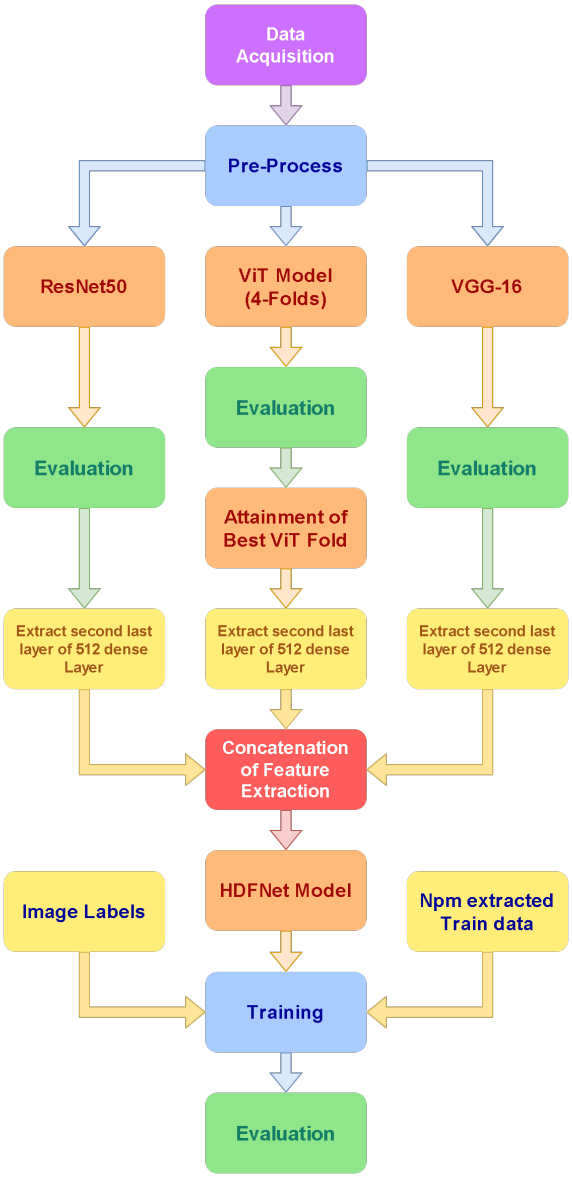


FIGURE 4: Flow diagram of the Proposed Method

be useful for ensuring the effectiveness and the speed of the learning model. Normalization is also applied to reduce the fluctuations between the pixel values into the range of 0-1 so that the learning model does not tend to focus on and develop a bias towards the more prominent features.

It has been generally noticed that when acquiring or dealing with medical datasets, certain categories or progression stages of diseases are rare enough to the extent that little data regarding those conditions exists. Therefore such datasets are typically found to be imbalanced with the majority class having the tendency to produce a bias in the machine learning architectures; this requires the usage of specialized techniques to address the problem and introduce a balance to the data. To balance the classes we have used High frequency balancing to reduce the over fitting, to some classes have numbers one fourth $1/4$ of the class which have the highest number of representation. We introduce augmentation, dropout layer , Batch Normalization reversed in a way that the minority class representation will equal to majorty class representation of samples. Therefore a highly specialized technique called High-frequency Balancing and Augmentation method is utilized to specifically introduce transformations into those spaces of the dataset where the minority classes appears with a higher frequency. Through ensuring that the minority class is appropriately represented in high frequency regions without damaging the overall integrity of the data, this technique easily allows for an enhancement to the generalization capabilities and the robustness of the AI model. For the final preparation, the overall data compiled for usage as input is divided into three subsets for usage during experimentation: 80% training data, 10% testing, and 10% validation.

V. CLASSIFICATION

Now that the data has been transformed into a suitable shape and adequately prepared to ensure better performance by the model, we introduce the different classification architectures utilized in this research and auxiliary techniques for further improvement in the performance of the models. Vision Transformer (ViT) is one of the models which we deployed for this experimentation [21] ; though unconventional, it is a great addition to the CNNs due to its unique capabilities of capturing global or long-range dependencies and context [22]. In our case, we considered a ViT (12-layer) which entails the division of the input dataset into 4 subsets, with all iterations of ViT using a different subset for validation whereas the remaining subsets would be used for testing. This approach allows the AI model to expand on the features captured from the data and allow for different perspectives to be formed which can then be combined for a more comprehensive feature map. [23] [24]

Some important components of the ViT that set it apart from other CNNs are the following: Patch Extractor, Token Embedder, Transformer Encoder, and the MLP. When the data is input to the ViT, the patch extractor divides each image into equal-sized patches upon which linear transformations are applied to transform them to a lower-dimensional vector space [25]. The Vision Transformer architecture consists of the following important layers: [26] [27]

- **Patch Extractor:** This component divides images into patches of size $P \times P$. Each image $x \in \mathbb{R}^{H \times W \times C}$ is split into $N = \frac{HW}{P^2}$ patches, where each patch is flattened into a vector $x_p \in \mathbb{R}^{P^2 \cdot C}$.

$$\{x_p^i\}_{i=1}^N \quad \text{where} \quad x_p^i \in \mathbb{R}^{P^2 \cdot C} \tag{1}$$

- **Token Embedder:** The flattened patches are then linearly transformed into a lower-dimensional vector space

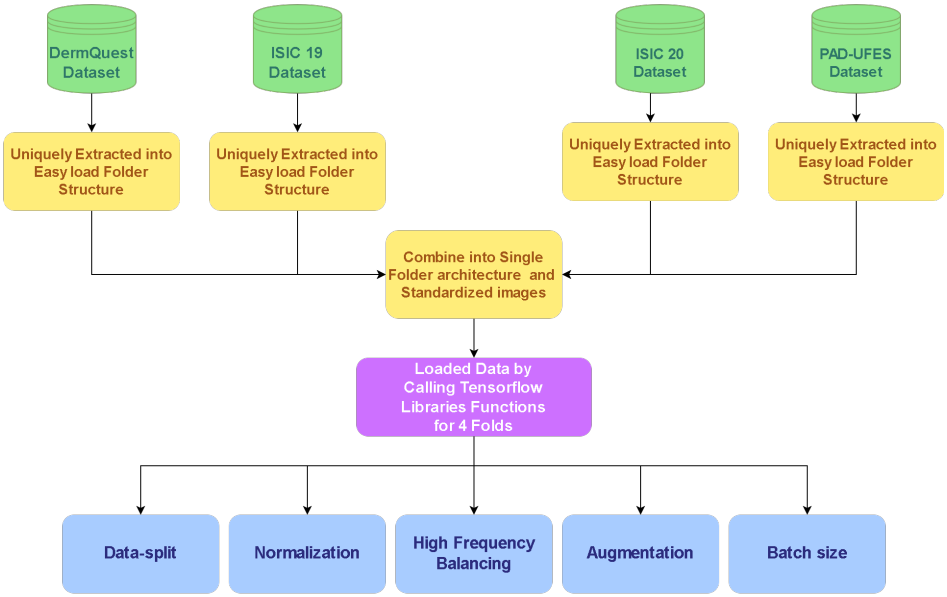


FIGURE 5: Proposed Preprocessing Method

using a learnable matrix $E \in \mathbb{R}^{(P^2 \cdot C) \times D}$, where D is the embedding dimension [28].

$$z_0 = [x_p^1 E; x_p^2 E; \ldots; x_p^N E] + E_{pos} \tag{2}$$

Here, E_{pos} is the positional encoding that adds positional information to the patch embeddings.

- **Transformer Encoder:** This component consists of L layers, each containing a Multi-Head Self-Attention (MSA) mechanism and a Feedforward Neural Network (FFN). The input to each layer is processed as follows:

$$z'_l = \text{MSA}(\text{LN}(z_{l-1})) + z_{l-1} \tag{3}$$

$$z_l = \text{FFN}(\text{LN}(z'_l)) + z'_l \tag{4}$$

where $l = 1, \ldots, L$, z_0 is the initial embedding, LN is layer normalization, and MSA is computed as:

$$\text{MSA}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h) W^O \tag{5}$$

with each attention head i defined as:

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{6}$$

and the scaled dot-product attention given by:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{7}$$

where W_i^Q, W_i^K, W_i^V, and W^O are learnable weight matrices.

- **MLP Head:** The final layer is a Multi-Layer Perceptron (MLP) that performs the classification. The MLP consists of multiple fully connected layers with non-linear activation functions such as GeLU or ReLU [29].

$$\text{MLP}(z_L) = W_2 \cdot \text{GeLU}(W_1 \cdot z_L + b_1) + b_2 \tag{8}$$

where z_L is the output from the last Transformer Encoder layer, and W_1, W_2, b_1, and b_2 are learnable parameters.

Other models utilized in this research include VGG-16 and ResNet-50.

A. VGG-16

VGG-16, comprising 13 convolutional layers and 3 fully connected layers, is renowned for its simplicity and uniform structure that uses 3×3 convolutional filters throughout the network, which is a relatively effective strategy for acquiring features. The key components include: [30] [31]

- **Convolutional Layers:** VGG-16 uses small 3×3 convolution filters with a stride of 1 and padding to preserve spatial resolution.

$$\text{Conv}(x) = W * x + b \tag{9}$$

where W is the filter, x is the input, and b is the bias [32].

- **ReLU Activation:** After each convolutional layer, a ReLU (Rectified Linear Unit) activation function is applied.

$$\text{ReLU}(x) = \max(0, x) \tag{10}$$

- **Max Pooling:** Pooling layers are used to reduce the spatial dimensions.

$$\text{MaxPool}(x) = \max(x_{i,j} \mid i, j \in \text{pooling window}) \tag{11}$$

- **Fully Connected Layers:** The output from the convolutional layers is flattened and passed through one or more fully connected layers.

$$\text{FC}(x) = Wx + b \tag{12}$$

where W and b are the weights and biases of the fully connected layer [33].

- **Softmax:** The final layer is a softmax layer used for classification.

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \tag{13}$$

where x_i is the input to the softmax function.

B. RESNET-50

ResNet-50 consists of 50 layers out of which 49 are convolutional and the final layer is fully connected. This machine learning architecture is distinguished by its usage of skip (or residual) connections which deal with the problem of vanishing gradient and allow for deeper networks to be prepared for advanced object detection and classification. The key components include:

- **Residual Block:** The core unit of ResNet-50 is the residual block, which includes a shortcut connection.

$$y = F(x, \{W_i\}) + x \tag{14}$$

where x is the input, $F(x, \{W_i\})$ is the residual function (e.g., a stack of convolutional layers), and y is the output.

- **Bottleneck Block:** ResNet-50 uses bottleneck blocks which include three layers: a 1×1 convolution, a 3×3 convolution, and another 1×1 convolution.

$$y = W_3\sigma(W_2\sigma(W_1x)) \tag{15}$$

where W_1, W_2, and W_3 are the weights of the convolutional layers, and σ is the ReLU activation function.

- **Identity and Convolutional Shortcuts:** Depending on whether the input and output dimensions match, identity shortcuts or convolutional shortcuts are used to adjust the dimensions [34].

$$y = F(x, \{W_i\}) + W_sx \tag{16}$$

where W_s is the weight matrix for the shortcut connection (if dimensions differ).

- **Global Average Pooling:** Before the final fully connected layer, a global average pooling layer is applied to reduce each feature map to a single value [35].

$$\text{GlobalAvgPool}(x) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_{i,j} \tag{17}$$

- **Fully Connected Layer:** The output from the global average pooling layer is passed through a fully connected layer for classification.

$$\text{FC}(x) = Wx + b \tag{18}$$

In the case of our ResNet-50 model consist of the following

- An initial convolutional layer and max-pooling layer.
- Four stages, each containing a stack of residual blocks.
- Each residual block has a series of convolutional layers with batch normalization and ReLU activation functions, followed by shortcut connections.

Residual Block: In there are two main types of residual blocks:

- Identity Block: Used when the input and output dimensions are the same.
- Convolutional Block: Used when the input and output dimensions are different, typically used to downsample the spatial dimensions.

Convolutional Block: A convolutional block in ResNet-50 also contains three convolutional layers, but with a slight difference:

- First Convolutional Layer: 1x1 convolution, reducing the dimensions.
- Second Convolutional Layer: 3x3 convolution, processing the data.
- Third Convolutional Layer: 1x1 convolution, restoring the dimensions.

Additionally, a convolutional block has a 1x1 convolution in the shortcut path to match the dimensions of the input and output.

The architecture of ResNet-50 can be broken down into the following stages:

- **Initial Layer**
  - 1 convolutional layer: 7x7, 64 filters, stride 2
  - 1 max-pooling layer: 3x3, stride 2
- **Stage 1**
  - 1 convolutional block: 1x1, 64 filters; 3x3, 64 filters; 1x1, 256 filters (plus a 1x1 convolution in the shortcut path).
  - 2 identity blocks: each with 1x1, 64 filters; 3x3, 64 filters; 1x1, 256 filters.
- **Stage 2**
  - 1 convolutional block: 1x1, 128 filters; 3x3, 128 filters; 1x1, 512 filters (plus a 1x1 convolution in the shortcut path).
  - 3 identity blocks: each with 1x1, 128 filters; 3x3, 128 filters; 1x1, 512 filters.
- **Stage 3**
  - 1 convolutional block: 1x1, 256 filters; 3x3, 256 filters; 1x1, 1024 filters (plus a 1x1 convolution in the shortcut path).
  - 5 identity blocks: each with 1x1, 256 filters; 3x3, 256 filters; 1x1, 1024 filters.
- **Stage 4**
  - 1 convolutional block: 1x1, 512 filters; 3x3, 512 filters; 1x1, 2048 filters (plus a 1x1 convolution in the shortcut path).
  - 2 identity blocks: each with 1x1, 512 filters; 3x3, 512 filters; 1x1, 2048 filters.
- **Final Layers**
  - 1 average pooling layer
  - 1 fully connected (dense) layer: 1000 units (for classification into 1000 classes in ImageNet).

Therefore, ResNet-50 has 49 convolutional layers, with the naming "ResNet-50" primarily referring to the depth of the network, including other types of layers as well [36] [37].

When the different models described above have processed the input and finished with their work, feature vectors are obtained through the extraction of their penultimate layers which are then fused through the concatenation feature fusion technique which allows for the dimensionality of the fused feature to be enhanced. In our case with 3 feature vectors of 512-dimensions each, our final feature vector would have a dimensionality of 1536. [24] [31] This feature fusion technique allows for the strengths of different models to be integrated into a single vector which when utilized as input for any model in the future allows the model to work with a greater efficiency. Now our compiled dataset and the fused feature vector are employed as input for a Deep Neural Network (DNN) where the feature vector is passed through a dense layer that reinforces the effect of the combined information for effective learning by a layer containing 1024 units which strengthens the convergence of the variety of feature representations onto a single vector.

These are all the details regarding our research methodology which clearly present our choices at different stages of the research. Next we shall move towards discussing the performance of the models and provide a conclusion to the research process.

VI. RESULTS AND DISCUSSION

It is pertinent to elaborate on the datasets that have been employed as input in this research before we move the reveal the results of our experimentation. Therefore the different datasets that we acquired are going to be briefly introduced now.

A. ISIC-2019

It is a treasure trove of useful dermoscopic images prepared for the ISIC's 2019 challenge that are 25,331 in number and are composed of various lesion categories that have crucial significance during research regarding medical imaging and skin cancer classification. The 8 categories of the dataset consist of Melanoma (MEL), Melanocytic Nevus (NV), Basal Cellular Carcinoma (BCC), Actinic Keratoses (AKIEC), Benign Keratosis-like Lesions (BKL), Dermatofibroma (DF), Vascular Lesions (VASC) and Squamous Cellular Carcinoma (SCC). Critical metadata has also been included such as age, sex, anatomic site along with precious contextual records for further insights and even a separate check dataset comprising the outlier class of data points not within the main data is present as well. The dataset also contains the images from ISIC-2018 (HAM10000) and ISIC-2017 challenges and can be easily sourced from the ISIC online archive or Kaggle.

B. ISIC-2020

The 2020 version of ISIC dataset contains 32,542 BKL-category images and 563 MEL images which while a mix of crucial skin conditions may constitute a class imbalance. Each file in this data contains patterns along with the visual photo, basic metadata and a unique patient identifier; the metadata elaborates on the minute details of the dataset thereby proving beneficial to dermatologists for research and experimentation. Histopathology has been employed to confirm each malignant diagnosis while a combination of expert assessment, long-term observation and histopathology has been employed for every benign case. Due to the existence of class imbalance in this dataset, the standard operation would be to employ augmentation or other specialized sampling methods to tackle the problem before being used as an input to the classifiers.

C. PAD-UFES

Containing a variety of medical images captured through the medium of smartphone devices, this dataset composed of 2298 rotoscope images taken from 1373 patients comprises 6 different lesion categories with 58.4% of the lesions being proven through biopsy, and is a data repository of great significance for skin cancer classification researches.

D. DERMQUEST-DERMIS

Offering a wide range of dermoscopic images accompanied by abundant data for research and analysis, this dataset caters to the needs of skin cancer classification researchers through the inclusion of some of the Melanoma (MEL) and NON-MEL categories of images. This is an excellent dataset that can be used as the foundation for projects which aim to deliver improved machine learning models with better accuracy and enhanced generalization.

E. LABEL DETAILS

In addition to the datsets that we have been elaborated upon here, some important labels regarding the categories of lesion images that we have used are also introduced here. There are four lesion categories that we have focused our research upon; BKL, MEL, NV, and BCC. BKL or Benign Keratosis-like Lesions are non-cancerous though difficult to differentiate from the cancerous lesions due to their resemblance. They look similar to warts and usually form due to long-term sun exposure for the adults. MEL or Melanoma is an aggressive, malignant category of lesions that affects melanocytes and is characterized by weird-shaped and colored moles. It forms as a result of ultraviolet exposure, sunburn, or genetic factors. NV or Nevi are benign melanocyte growths which are typically referred to as moles that may be innate or acquired later in life. They may be flat or raised and their colors are limited to brown or black though the irregular borders and the different colors may lead to them being confused with melanoma. Finally, the BCC or Basal Cell Carcinoma is the least aggressive and the most common form of skin cancer which while, unable to metastasize, can still wreak havoc on its locality if untreated. This form of lesion may appear veiny with flesh or brownish colors and usually forms because of exposure to carcinogens, radiation, chronic sun, or even something as simple as older age.

TABLE 1: Summary of the images and classes in the acquired datasets

No.	Dataset	Total Categories	Short Name of Categories	Total Images
1	ISIC 2019 Dataset	8	AKL, BCC, MEL SCC, BKL, NV, DF, VASC	25,331
2	PDF-UFES	6	ACK, BCC, MEL, SCC, NEV, SEK	1,612
3	ISIC 2020 Dataset	2	BKL, MEL	33,126
4	DERM QUEST DERMIS Dataset	2	MEL, NOTMEL	180

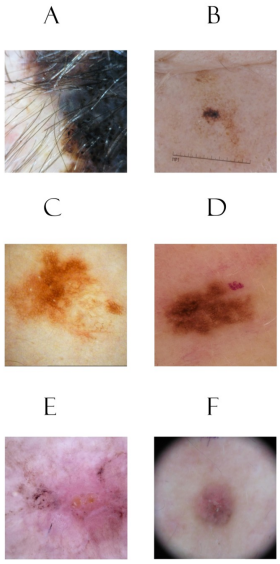


FIGURE 6: Sample images of dataset A). MEL class, B) NV class, C,D) BKL class, E,F) BCC class.

F. METRIC FOR EVALUATION

Finished with the description of the datasets and the label details, we now proceed to the description of our selection of evaluation metrics. We chose to employ the AUC-ROC score, Recall, and F1-Score to attain a comprehensive understanding of our models' performance.

1) Receiver Operation Curve (ROC)

ROC is a graphical representation that plots the curve between the true positive rate and the false positive rate. It is derived through the calculation of specificity and sensitivity.

$$\text{ROC} = \{(\text{Specificity}_i, \text{Precision}_i)\}_{i=0}^{n}$$

2) Specificity

Specificity measures the proportion of true negatives that are correctly identified by the neural network architecture. It is calculated through the following formula:

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{19}$$

Where:

- FP = False Positives
- TN = True Negatives

3) Precision

Precision refers to the proportion of true positives that are correctly identified by the neural network. Its calculation is carried out through the following process:

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}} \tag{20}$$

Where:

- TP = True Positives
- FP = False Positives

4) AUC

Area Under the ROC Curve or AUC refers to a numerical value existing between 0 and 1 which signifies the efficiency levels of the classifier architectures after calculating the ability of the classifier to correctly identify positive and negative classes.

$$\text{AUC} = \int_0^n \text{Sensitivity}(\text{Specificity})\, d\text{Precision} \tag{21}$$

In the domain of medicine where different categories or progression stages of the disease may vary in terms of their rarity, it is imperative to utilize a metric that has the ability to display varied discriminating thresholds to account for this variation. In contrast simple accuracy has a simple fixed threshold without any flexibility and is therefore not suitable for usage as our metric, compared to the AUC-ROC. Furthermore two crucial features of the diagnosis are processed by the AUC-ROC: true positives and true negatives, allowing professionals to easily evade the dangers of false positives or false negatives which could constitute fatal consequences in real-world settings. In addition, our metric also utilizes the moving threshold to deal with the issue of class imbalance and is therefore a more reliable performance evaluator than general accuracy.

5) Recall

Recall is another machine learning evaluation metric used to define how often the employed model calculates and correctly identifies the class positive instances. Following is the formula of recall function:

$$R = \frac{TP}{TP + FN} \tag{22}$$

6) F1 Score

F1 score is a machine learning evaluation metric is generally used to evaluate how many times model has correctly evaluated correct prediction across data set using precision and recall. Following Formula is used to calculate the F1-score:

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{23}$$

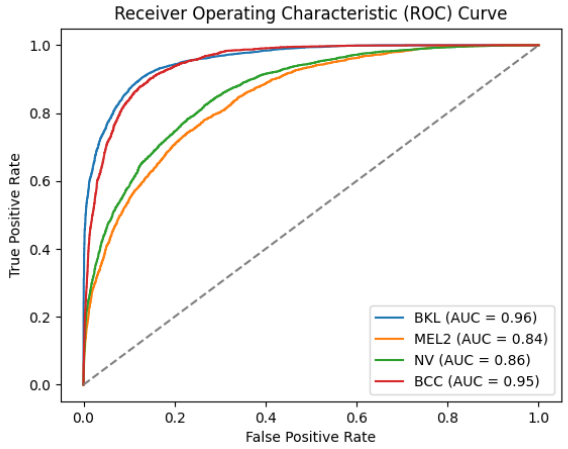


FIGURE 7: AUC-ROC curve of Proposed Custom Hybrid Net Dense Fusion Model.

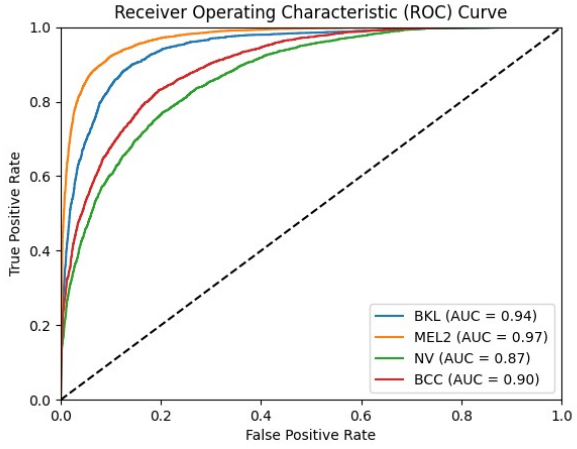


FIGURE 8: AUC-ROC curve achieved using ViT.

Based on the plotted graphs displayed above, it is easily noticeable that the Vision Transformer model's performance pales in comparison to the feature fused DNN. The former's performance scores for the different lesion categories are 0.96 (BKL), 0.84 (MEL2), 0.86 (NV) and 0.95 (BCC) where as the latter's scores are 0.94 (BKL), 0.97 (MEL2), 0.87 (NV) and 0.90 (BCC). In terms of class discrimination capabilities it is evident that the DNN performs better in terms of MEL and NV compared to BKL and BCCAs is noticeable. Further information regarding F1-score, recall and precision is depicted in the diagram below.

G. EVALUATION

Based on the information above, the precision, recall and f1-scores for the different classes of lesions are 0.95, 0.94 and 0.95 for BKL, 0.97, 0.96 and 0.96 for MEL, 0.88, 0.89 and 0.88 for NV, and 0.91, 0.90 and 0.90 for BCC, respectively. Furthermore we can easily notice the overall performance differences between the previous models and the feature fused DNN. While ViT, ResNet50 and VGG16 achieved an overall AUC-ROC score of 90.25%, 90.27% and 89.75% respectively, the feature fused DNN shows a significant improvement through its score of 92.15%.

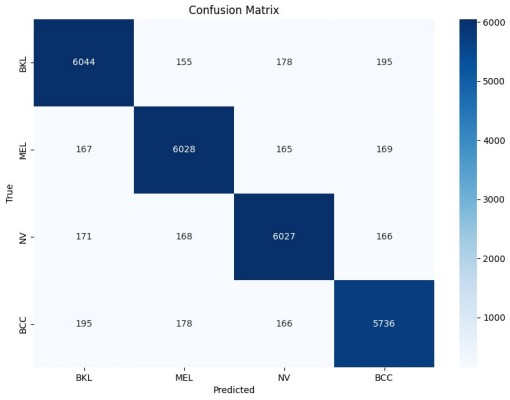


FIGURE 9: Confusion Matrix of Proposed Model.

The Above Figure illustrated presents the performance of Custom CNN model attained by averaging fusion and illustrates the True Positives, False Positive, True Negatives and False Negative prediction of labels for each label class in the study

The significant improvement in results that we see here results from factors such the contribution of the strengths of different classifiers through the unified feature map which allows the DNN to be comprehensively strengthened when dealing with the dataset thus leading to a more comprehensive analysis of the features which leads to better generalization and greater efficiency. For future work, the feasibility of inclusion of other CNN architectures can be tested by the researchers along with addressing the lack of external validity. In addition, the future research could extend by expanding on the four categories of lesion that we have experimented on and include a greater number of categories to work with. Finally, the application of this methodology in real-time settings can also be tested in the future.

VII. CONCLUSION AND FUTURE WORK

We conclude our research here with the Deep Neural Network being provided with a unified feature vector that was obtained

TABLE 2: Precision, Recall F1-core, AUC-ROC and overall Accuracy performance on the various model of vgg, Resenet Series, ViT model and proposed fusion model.

Models	Precision				Recall				F1-Score				AUC-ROC	Overall Accuracy
	BKL	MEL	NV	BCC	BKL	MEL	NV	BCC	BKL	MEL	NV	BCC		
VGG16	0.87	0.80	0.79	0.68	0.77	0.87	0.76	0.70	0.76	0.77	0.76	0.70	0.8975	0.78
VGG22	0.81	0.75	0.72	0.76	0.75	0.55	0.80	0.81	0.78	0.68	0.73	0.88	0.88	0.73
ResNet50	0.82	0.75	0.73	0.85	0.74	0.65	0.80	0.80	0.79	0.70	0.73	0.76	0.89	0.75
ResNet101	0.81	0.74	0.75	0.86	0.74	0.66	0.81	0.80	0.79	0.71	0.76	0.76	0.895	0.75
4 Fold ViT	0.90	0.91	0.89	0.90	0.90	0.90	0.88	0.91	0.92	0.90	0.88	0.91	0.9025	0.90
Feature Fused Model	0.95	0.97	0.89	0.91	0.94	0.96	0.88	0.90	0.92	0.95	0.88	0.90	0.9215	0.92

TABLE 3: Comparison of performance of the EfficientNet Series on both AUC-ROC and Accuracy

Model	EfficientNetB1	EfficientNetB2	EfficientNetB3	EfficientNetB4	EfficientNetB5	EfficientNetB6
AUC-ROC	0.86	0.86	0.82	0.85	0.88	0.87
Accuracy	0.77	0.77	0.78	0.77	0.80	0.81

TABLE 4: Comparison of Past performed researches with the Proposed Models with applied Datasets

Reference	Year	Dataset	Accuracy / AUC
Kahia et al [17]	2022	Melanoma images: 374 Nevus images: 1372 Seborrheic keratosis images: 254	Accuracy = 73.33%
Rezaoana et al [13]	2021	Kaggle competition 2019 Images=25,780	Accuracy = 79.45%
Kaya et al [2]	2023	Kaggle competition 2021 Total images used=3297	Accuracy = 83.33%
Ali et al [18]	2022	HAM10000 dataset 10015 images	Accuracy = 87.91
Jaisakthi et al [11]	2022	ISIC 2019 and ISIC 2020 skin cancer classification. ISIC 2020=33,126 ISIC 2019=25331 It includes BCN_20000, HAM10000, and MSK	AUC-ROC = 91%
Gouda et al [5]	2022	ISIC 2018 dataset Total images = 11,527	AUC-ROC = 85.8%
Bassel et al [3]	2021	ISIC 2019 Challenge Used 1800 images	AUC-ROC = 90.9%
Nakai et al [10]	2022	ISIC2017 =2750 Three Classes Dataset	AUC-ROC = 92.1%
Akter et al [1]	2022	HAM10000 dataset 10015 images	AUC-ROC = 78%
Jain et al [6]	2021	HAM10000 dataset 10015 images	AUC-ROC = 90.48%
Proposed VIT Model	**2024**	**ISIC 2019, ISIC 2020, PDF-UFES, and DERMQUEST-DERMIS Dataset**	**Acc = 90%, AUC-ROC = 90.25%**
Proposed Hybrid Net Dense Fusion Model	**2024**	**ISIC 2019, ISIC 2020, PDF-UFES, and DERMQUEST-DERMIS Dataset**	**Acc = 92%, AUC-ROC = 92.15%**

from concatenation of the corresponding feature extractions acquired from ResNet-50, VGG-16, and ViT (12-layer), along with the labels from the validation set that resulted in us acquiring an AUC-ROC score of 92.15%. The respective AUC-ROC scores of the different categories selected for this research are BKL (0.94), BCC (0.90), NV (0.87), and MEL (0.97). This level of efficiency is unprecedented among the different models included at any stage of this research. When testing ResNet-50, VGG-16, and ViT for the classification task so as to establish a comparative point to emphasize the significance of the DNN's results, we discovered their AUC-ROC scores to be hovering around 0.90 or 90%.

Throughout the course of this research, a number of models that had been considered initially were left behind in the end as well due to a variety of reasons. For example, VGG-22 was dropped due to the Vanishing Gradient problem while the choice of hyper parameters induced a tendency to shift the overall nature of classification towards overfitting; under these circumstances, the results attained, even if good, would be biased and horribly imbalanced. To deal with this overfitting problem would have required extensive hyper parameter tuning to prepare VGG-22 for inclusion into the research. Another model dropped from this article was the ResNet-101 whose results of 89% were not considered excellent enough to be included in the research and would have lowered the AUC-ROC of the DNN, if included. In the case of feature fusion implication the reason we didn't implement the employement is due to the fact V-22 and resnet-101, efficient were drop out because feature extraction process was same as vgg-16 and resnet-50 so they may not be very practical.

This research leaves quite a few aspects open for further experimentation and research. First of all, future researches should try to carry out a comparative analysis of the performance of other types of CNN architectures by applying the same methodology so that the feasibility and versatility of this methodology can be assessed. Secondly future researches should address the problem of external validity as this repetitive process is used to determine the generalization

capabilities of the model through unseen data. Additionally, while our research was limited to the four specific categories of skin cancer lesions therefore future research can proceed further by focusing on a greater number of categories. Finally, there are great prospects in the field of applied research where the practical effectiveness of the methodology is tested under real-time in an actual medical setting.

REFERENCES

[1] M. S. Akter, H. Shahriar, S. Sneha, and A. Cuzzocrea, "Multi-class skin cancer classification architecture based on deep convolutional neural network," in *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 2022, pp. 5404–5413.

[2] V. Kaya and İ. Akgül, "Classification of skin cancer using vggnet model structures," *Gümüşhane Üniversitesi Fen Bilimleri Dergisi*, vol. 13, no. 1, pp. 190–198, 2022.

[3] A. Bassel, A. B. Abdulkareem, Z. A. A. Alyasseri, N. S. Sani, and H. J. Mohammed, "Automatic malignant and benign skin cancer classification using a hybrid deep learning approach," *Diagnostics*, vol. 12, no. 10, p. 2472, 2022.

[4] S. K. Datta, M. A. Shaikh, S. N. Srihari, and M. Gao, "Soft attention improves skin cancer classification performance," in *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data: 4th International Workshop, iMIMIC 2021, and 1st International Workshop, TDA4MedicalData 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 4*. Springer, 2021, pp. 13–23.

[5] W. Gouda, N. U. Sama, G. Al-Waakid, M. Humayun, and N. Z. Jhanjhi, "Detection of skin cancer based on skin lesion images using deep learning," in *Healthcare*, vol. 10, no. 7. MDPI, 2022, p. 1183.

[6] S. Jain, U. Singhania, B. Tripathy, E. A. Nasr, M. K. Aboudaif, and A. K. Kamrani, "Deep learning-based transfer learning for classification of skin cancer," *Sensors*, vol. 21, no. 23, p. 8142, 2021.

[7] 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. 114 822–114 832, 2020.

[8] Z. Li, Z. Chen, X. Che, Y. Wu, D. Huang, H. Ma, and Y. Dong, "A classification method for multi-class skin damage images combining quantum computing and inception-resnet-v1," *Frontiers in Physics*, vol. 10, p. 1046314, 2022.

[9] S. B. Mukadam and H. Y. Patil, "Skin cancer classification framework using enhanced super resolution generative adversarial network and custom convolutional neural network," *Applied Sciences*, vol. 13, no. 2, p. 1210, 2023.

[10] K. Nakai, Y.-W. Chen, and X.-H. Han, "Enhanced deep bottleneck transformer model for skin lesion classification," *Biomedical Signal Processing and Control*, vol. 78, p. 103997, 2022.

[11] J. SM, M. P, C. Aravindan, and R. Appavu, "Classification of skin cancer from dermoscopic images using deep neural network architectures," *Multimedia Tools and Applications*, vol. 82, no. 10, pp. 15 763–15 778, 2023.

[12] A. Al-Rasheed, A. Ksibi, M. Ayadi, A. I. Alzahrani, and M. Mamun Elahi, "An ensemble of transfer learning models for the prediction of skin lesions with conditional generative adversarial networks," *Contrast Media & Molecular Imaging*, vol. 2023, pp. 1–15, 2023.

[13] N. Rezaoana, M. S. Hossain, and K. Andersson, "Detection and classification of skin cancer by using a parallel cnn model," in *2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*. IEEE, 2020, pp. 380–386.

[14] M. R. Thanka, E. B. Edwin, V. Ebenezer, K. M. Sagayam, B. J. Reddy, H. Günerhan, and H. Emadifar, "A hybrid approach for melanoma classification using ensemble machine learning techniques with deep transfer learning," *Computer Methods and Programs in Biomedicine Update*, vol. 3, p. 100103, 2023.

[15] G. Yang, S. Luo, and P. Greer, "A novel vision transformer model for skin cancer classification," *Neural Processing Letters*, vol. 55, no. 7, pp. 9335–9351, 2023.

[16] Z. Zhao, "Skin cancer classification based on convolutional neural networks and vision transformers," in *Journal of Physics: Conference Series*, vol. 2405, no. 1. IOP Publishing, 2022, p. 012037.

[17] M. Kahia, A. Echtioui, F. Kallel, and A. B. Hamida, "Skin cancer classification using deep learning models." in *ICAART (1)*, 2022, pp. 554–559.

[18] K. Ali, Z. A. Shaikh, A. A. Khan, and A. A. Laghari, "Multiclass skin cancer classification using efficientnets–a first step towards preventing skin cancer," *Neuroscience Informatics*, vol. 2, no. 4, p. 100034, 2022.

[19] S. Medhat, H. Abdel-Galil, A. E. Aboutabl, and H. Saleh, "Skin cancer diagnosis using convolutional neural networks for smartphone images: A comparative study," *Journal of Radiation Research and Applied Sciences*, vol. 15, no. 1, pp. 262–267, 2022.

[20] P. Banasode, M. Patil, and N. Ammanagi, "A melanoma skin cancer detection using machine learning technique: support vector machine," in *IOP Conference Series: Materials Science and Engineering*, vol. 1065, no. 1. IOP Publishing, 2021, p. 012039.

[21] P. Zhang, X. Dai, J. Yang, B. Xiao, L. Yuan, L. Zhang, and J. Gao, "Multi-scale vision longformer: A new vision transformer for high-resolution image encoding," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 2998–3008.

[22] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao, "Pre-trained image processing transformer," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 12 299–12 310.

[23] J. Chen, Y. He, E. C. Frey, Y. Li, and Y. Du, "Vit-v-net: Vision transformer for unsupervised volumetric medical image registration," *arXiv preprint arXiv:2104.06468*, 2021.

[24] I. Ali, M. Muzammil, I. U. Haq, M. Amir, and S. Abdullah, "Deep feature selection and decision level fusion for lungs nodule classification," *IEEE Access*, vol. 9, pp. 18 962–18 973, 2021.

[25] D. Zhou, B. Kang, X. Jin, L. Yang, X. Lian, Z. Jiang, Q. Hou, and J. Feng, "Deepvit: Towards deeper vision transformer," *arXiv preprint arXiv:2103.11886*, 2021.

[26] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10 012–10 022.

[27] S. H. Lee, S. Lee, and B. C. Song, "Vision transformer for small-size datasets," *arXiv preprint arXiv:2112.13492*, 2021.

[28] Y. Wang, R. Huang, S. Song, Z. Huang, and G. Huang, "Not all images are worth 16x16 words: Dynamic transformers for efficient image recognition," *Advances in neural information processing systems*, vol. 34, pp. 11 960–11 973, 2021.

[29] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, "Pyramid vision transformer: A versatile backbone for dense prediction without convolutions," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 568–578.

[30] L.-Y. Ye, X.-Y. Miao, W.-S. Cai, and W.-J. Xu, "Medical image diagnosis of prostate tumor based on psp-net+ vgg16 deep learning network," *Computer Methods and Programs in Biomedicine*, vol. 221, p. 106770, 2022.

[31] X. Cheng, L. Tan, and F. Ming, "Feature fusion based on convolutional neural network for breast cancer auxiliary diagnosis," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–10, 2021.

[32] S. A. Y. Al-Galal, I. F. T. Alshaikhli, M. Abdulrazzaq, and R. Hassan, "Brain tumor mri medical images classification with data augmentation by transfer learning of vgg 16," *Journal of Engineering Science and Technology*, pp. 21–32, 2021.

[33] D. Albashish, R. Al-Sayyed, A. Abdullah, M. H. Ryalat, and N. A. Almansour, "Deep cnn model based on vgg16 for breast cancer classification," in *2021 International conference on information technology (ICIT)*. IEEE, 2021, pp. 805–810.

[34] A. Sharma and P. Dutta, "The computational approach for brain hemorrhage classification using deep transfer learning," *Available at SSRN 4243426*, 2022.

[35] Y. S. Devi and S. P. Kumar, "A deep transfer learning approach for identification of diabetic retinopathy using data augmentation," *International Journal of Artificial Intelligence*, vol. 11, pp. 1287–1296, 2022.

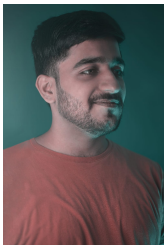
[36] H. Fauzi, R. B. Ansori, T. Siadari, A. B. Harsono, and Q. N. Rahmah, "Classification of cervical cancer images using deep residual network architecture," *International Journal of Artificial Intelligence Research*, vol. 7, no. 1, pp. 56–63, 2023.

[37] Y. Yu, H. Lin, J. Meng, X. Wei, H. Guo, and Z. Zhao, "Deep transfer learning for modality classification of medical images," *Information*, vol. 8, no. 3, p. 91, 2017.

SYED NEHAL HASSAN SHAH received the M.S. degree in Artificial Intelligence from Air University, Islamabad, Pakistan, in 2024. Since 2023, he has been working as an AI Developer at Octaloop Technologies. His research interests include deep learning, computer vision, and medical image analysis, with a particular focus on vision transformers. Mr. Shah has made significant contributions to facial recognition systems, notably improving an Nvidia Deepstream-based pipeline to process 25 live streams concurrently and optimizing the Arcface facial recognition model. He also developed the HybridNet Dense Fusion Model for multi-class cancer detection. His current projects involve the multistage classification of ulcerative colitis using colonoscopy images and skin lesion cancer detection using dermoscopic images.

IMRAN TAJ is currently working as an Assistant Professor at Zayed University. Prior to that he was a Senior Team Lead, Information Systems Branch with BC Public Service, Canada, where he led the digital transformation and automation of an enterprise software application using Artificial Intelligence techniques and Robotic Process Automation tools. He received his Ph.D. degree in computer engineering from University of Paris-Est, France. He has several years of professional experience of applying machine learning operations to craft real world data solutions and has been involved in all phases of Software Development Life Cycle - from inception to implementation - for several systems engineering projects in the fields of Artificial Intelligence, Machine Learning and Data Sciences.

SYED MUHAMMAD USMAN (Member, IEEE) is working as Assistant Professor in Air University, Islamabad, Pakistan. He completed Phd in Computer Engineering from Bahria University, Islamabad and MS in Computer Engineering from NUST, Islamabad, Pakistan. He has teaching and research experience of more than eight years with research interests in biomedical signal processing, medical imaging and precision agriculture. He has more than 20 publications in international journals and peer reviewed conferences. He is also reviewer of IEEE and Elsevier journals.



SYED ABDULLAH SHAH is a dedicated student poised to complete his Master's in Artificial Intelligence from Air University, Islamabad, Pakistan. With a robust academic foundation, he holds a Bachelor's degree from Riphah International University, I-14 Islamabad. He a Pakistani citizen hailing from Punjab born in 25th January 2000. Mr. Abdullah has demonstrated his scholarly aptitude through his involvement in two recent to be be published research papers. Currently focused on his studies, Syed continues to explore the vast potential of artificial intelligence, aiming to make impactful advancements in this rapidly evolving domain.

ALI SHARIQ IMRAN (Member, IEEE) received a master's degree in software engineering and computing from the National University of Science and Technology (NUST), Pakistan, in 2008 and a Ph.D. in computer science from the University of Oslo (UiO), Norway, in 2013. He is associated as an Associate Professor with the Department of Computer Science (IDI), Norwegian University of Science and Technology (NTNU), Norway. With over 15 years of teaching and research experience, he devised innovative ways to design effective multimedia learning objects and integrate the teaching-research nexus frameworks at the graduate level. He served as a commission member of the Ministry of Education of Macedonia in setting up Mother Theresa University in Skjope. He leads a capacity-building project called CONNECT (https://norpart-connect.com) funded by the Higher Education Commission of Norway, DIKU, under the NORPART scheme as a coordinator and three Erasmus+ KA2 projects (PhDICTKES (https://phdictkes.eu), RAPID, and TKAEDiT) as a project manager at NTNU, along with an Excited mini-project funded by NTNU. Dr. Ali also leads a research group on Deep NLP (http://deep-nlp.net) and specializes in applied deep learning research to address various multi-modality media analysis application areas for audio-visual and text processing. He has co-authored over 100 peer-reviewed journals and conference publications and has served as an editor and reviewer for many reputed journals. He is a member of the Intelligent Systems and Analytics research group at NTNU and an IEEE/ACM Member.

SHEHZAD KHALID is a passionate academician, researcher and research management professional with 22+ years of experience and a proven track record of delivering high-quality results in the field of Machine Learning, Computer vision, Signal Processing, Natural Language Processing, Intelligent Medical Diagnostics etc. Possesses a strong background in executive management of Postgraduate Programs, Research Innovation Commercialization (RIC) and Entrepreneurial Ecosystem in academic organization with focus on conducive policy development and execution. Led RIC ecosystem and Incubation center at Bahria university, which has been recognized amongst top ranked ecosystems in Pakistan. Proficient in leveraging advanced research techniques in different applied domains as reflected by track record of around 100 high quality journal publications, 50 conference publications, tens of PhD/MPhil supervisions and successful completion of Rs. 65 million+ worth of research and entrepreneurial projects.