Journal of Knowledge Learning and Science Technology
ISSN: 2959-6386 (Online)
2024, Vol. 3, No. 4, pp. 169–187
DOI: https://doi.org/10.60087/jklst.v3.n4.p169

Research Article

Skin disease classification using two path deep transfer learning models

Ram Chandna[1], Aarav Bansal[2], Aryan Kumar[3], Shrestha Hardia[4], Omogbolaga Daramola[5], Anirudh Sahu[6], Kritika Verma[7], Karan Dhingra[8], Saloni Verma[9]

[1]Independent Researcher, Uttarakhand, India
[2]Independent Researcher, Uttar Pradesh, India
[3]Independent Researcher, Uttar Pradesh, India
[4]Independent Researcher, Madhya Pradesh, India
[5]Independent Researcher, Lagos, Nigeria
[6]Independent Researcher, Madhya Pradesh, India
[7]Syracuse University, Department of Computer Science, New York, United States of America
[8]University of Ottawa, Department of Biomedical Engineering, Ontario, Canada
[9]Cornell University, Department of Biomedical Engineering, New York, United States of America

Abstract

Skin diseases are among the most common diseases that affect millions of lives per year, yet diagnosing these has several complexities even for trained dermatologists due to overlapping symptoms and features in several diseases. A myriad of deep learning models have been proposed as a solution for diagnosing but a clinically useful model with high accuracy multi disease classification and lower computational complexity is still unavailable. This study focuses on comparing different image pre-processing techniques, transfer learning models and ensemble learning techniques to build a computationally cheap model for 8-class identification of skin diseases. A two path model with EfficientNet and MobileNetV2 transfer learning models as base feature extractors and a final model that stacks the two model results and classifies the images into one of the eight classes is used. The model is trained and tested on ISIC-2019 dataset for 8 class image classification that involves the three types of skin cancers as well. The dataset has an extreme class imbalance problem which leads to favored prediction of the classes with more image, for this first we run simple image augmentation. Secondly, two distinctly processed images are created from each initial image. The two path model takes the two images, gives each to a base model and combines the two outputs, enabling the classifier to consider different features that become prominent due to dissimilar preprocessing techniques. The model is tested with new images on multiple standard metrics to get a final overview of its performance, it gives a diagnosing accuracy of 70% which is

[*]Corresponding author: Ram Chandna

Email addresses:

ramchandna3@gmail.com (Ram Chandna), aarav2008bansal@gmail.com (Aarav Bansal), aryanpsmail@gmail.com (Aryan Kumar), hardiashrestha@gmail.com (Shrestha Hardia), gbolagadaramola765@gmail.com (Omogbolaga Daramola), renudilip19@gmail.com (Anirudh Sahu), krverma@syr.edu (Kritika Verma), k2dhingr@gmail.com (Karan Dhingra), sv458@cornell.edu (Saloni Verma)

Received: DD MM 2024; **Accepted:** DD MM 2024; **Published:** DD MM 2024

close to some state of the art models that consume higher computational power. The classification results imply that by further improving the data gathering and preprocessing techniques along with exploring other base transfer learning models the results of the final model can be reliable while maintaining a low computational requirement, making the diagnoses accessible. This also highlights that such two path algorithms that employ simpler models could be useful for multi class classification tasks where differently processed images might be required to extract features of distinct diseases.

Keywords

Machine Learning, Deep Learning, Convoluted Neural Networks, Transfer Learning, Image Classification, Disease Diagnosis

1. Introduction

Skin disease impacts millions of individuals being one of the most common health issues globally with skin cancer being most severe. Skin cancer is a form of cancer for which around 1.5 million new cases were diagnosed in 2020 including about 330,000 instances of Melanoma, a more severe type and continuing to increase even today. Among these cases roughly 57,000 are fatal [1][2]. Along with Melanoma there are Non Melanoma skin cancers like Basal Cell Carcinoma (BCC) and Squamous Cell Carcinoma (SCC) are more common in countries like the US, where 3.6 million BCC cases and 1.8 million SCC cases are diagnosed annually. These cases occur in about 3.3 million people combined as some individuals have more than one lesion treatment per year [3]. While non melanoma skin cancers can be treated effectively and are typically non life threatening they may cause discomfort and anxiety due to side effects such as social limitations and physical complications that can diminish overall quality of life [4]. Moreover the financial burden associated with treating and monitoring these conditions is substantial for both patients and the economy at large. The average annual cost per person for treating melanoma skin cancer was $1,200, in 2018 while melanoma treatment costs reached $2,400 [5]. The yearly medical expenses for managing skin cancer in the United States amount to $8.9 billion [6]. Additionally, timely identification of skin cancer plays a role in lessening the severity and mortality rates of these conditions, especially melanoma, which boasts a 99 percent survival rate over a five year period [3].

Dermoscopy is a non-invasive imaging technique that allows dermatologists to closely examine skin lesions with enhanced magnification [7]. This method provides a detailed view that goes beyond what can be seen with the naked eye which makes it essential for early diagnosis of skin conditions, as accurate evaluation of the disease by visual inspection only requires extensive training and experience. There have been numerous microfluidic based approaches used for pathogen detection [8] however, the visual similarity between different types of skin cancer lesions often complicates the diagnostic process, even for highly experienced medical practitioners [9].

Advancement in image preprocessing and machine learning have revolutionized the diagnosis and treatment of skin diseases, leading to the development of faster and more precise diagnostic methods. In medical image analysis, deep learning (DL) algorithms are now the preferred tools for visual based disease classification. The efficiency and accuracy with which these algorithms analyze large datasets and extract relevant features from image data are superior to traditional machine learning (ML) algorithms [10]. Moreover, traditional ML algorithms such as Support Vector Machine (SVM), K Nearest Neighbor (KNN), Decision Tree, require manually extracted features [11] which aren't generally ideal for medical image diagnosis as new/complex variations of diseases are diagnosed from time to time. The Convolutional Neural Network (CNN), in particular, is a powerful deep learning architecture used for image processing and classification. Its ability to learn and extract useful characteristics from raw image data automatically without human supervision makes it a default option for these tasks [12]. Modifications to the CNN architecture have given rise to a variety of CNN-based models such as Inception-v4 [13], VGGNet [14], ResNeXt [15], MobileNetV2 [16], DenseNet [17], EfficientNet [18] with each model possessing unique advantages over the other depending on variable requirements and computational costs. In combination with other deep learning algorithms such as LSTMs [19] [20], vision transformers [21] or autoencoders, the CNNs' performance on medical image classification tasks is further enhanced. Several research studies have employed CNNs in dermatological image classification, achieving notable model performance [22].

Research in 2017, trained a deep CNN, utilizing the Inception v3 CNN architecture, on a dataset containing 129,450 skin disease images. A tree-structure taxonomy of the disease class, organized based on the visual and clinical similarities of the disease, was introduced. The model achieved higher accuracy than expert dermatologists for a broader 9 class classification [23]. In 2018, The diagnostic accuracy of a deep learning model in melanoma detection was tested via a comparison of its performance with a large group of 58 dermatologists by Haenssle et al [24]. The CNN-model achieved a greater ROC-AUC of 86% using only dermoscopic images, compared to 82% mean ROC-AUC of the dermatologists, who used both dermoscopic and clinical information. Another research proposed a model that achieved a 92.90% accuracy by using a combination of CNNs and a one-versus-all approach. In this approach, seven different models are trained, one for each of the seven skin diseases that the researchers aimed to identify. Each model is trained to differentiate a particular disease from the other six diseases combined. This means that each model essentially learns a specific disease's features and classifies an image on a binary basis i.e. whether it contains those features or not. Finally, the image is assigned the disease class predicted by the model with the highest true output [25]. In this research we used two Transfer Learning models as feature extractors and concatenated the two outputs to pass through a final fully connected classification layer. By utilizing the feature extraction knowledge acquired from training on the state-of-the-art ImageNet dataset, models such as MobileNet and Efficient Net demonstrate good performance in classifying dermatological images after fine tuning on the ISIC-2019 dataset for a 8 class skin disease classification. Additionally, combining predictions from multiple models through ensembling techniques enhances accuracy, which highlights adaptability of deep learning in managing intricate classification challenges. Instead of creating a single preprocessed image, the two paths of the ensembling model i.e. the MobileNet path and the EfficientNet path get two differently processed images one where the lesion area is segmented with the help of pixel values of different color intensities and one where only filtration and color inversion is applied. This dual path system helps to use a combination of preprocessing techniques which help in distinct feature extraction that are highlighted through dissimilar data processing methods.

2. Discussion

2.1. Datasets for Dermatological Diseases

The datasets considered for this research included the following.

2.2.1. Dermnet

Freely available large dataset of over 23,000 images with 23 super-classes [26]. It is maintained by a diverse group of dermatologists and images are from various patients and locations. The major disadvantages associated with it are that unlike ISIC challenge datasets, it lacks a separate test dataset and it might lead to inconsistent results when researchers split the dataset by themselves [27].

2.2.2. PH2

Publicly available dataset made in collaboration with University of Porto, University of London and Dermatology Service of Hospital Pedro Hispano. It is a detailed high-quality annotated dataset which contains manual segmentation of skin lesions, clinical diagnosis and identification of features on the basis of asymmetry, colors, pigment network, dots, streaks, and regression areas. The disadvantage with it is that it only includes 200 images and addition of more images require manual annotation effort which might be time and effort intensive [28].

2.2.3. HAM 10000

10,015 dermatoscopic images collected over a period of 20 years by the Department of Dermatology at the Medical University of Vienna, Austria and skin cancer practice of Cliff Rosendahl in Queensland, Australia. The issue with this dataset lies in complexity and sophisticated algorithms required to fully exploit its full capabilities [29].

2.2.4. ISIC 2019 Challenge Dataset

Largest collection of dermatoscopic images available with 25,331 images classified into eight different classes - Melanoma (MEL), Dermatofibroma (DF), Melanocytic Nevus (NV), Basal Cell Carcinoma (BCC), Benign Keratosis (BKL), Vascular Lesion (VASC), Actinic Keratosis (AK), Squamous Cell Carcinoma (SCC). It provides a good benchmark and industry standard for evaluating different models [29, 30].

2.2.5. Selection Process

A common issue prevalent over almost every dataset including the ISIC 2019 and ISIC 2020 was class imbalance. Class imbalance refers to unequal distribution of data into different classes which might result in overfitting or underfitting among models. For example: 6705 images for Melanocytic nevi (NV) compared to only 115 images of Dermatofibroma (DF) in the HAM10000 dataset [29]; or 12,000 images for Melanocytic Nevus (NV) compared to only

239 images for Dermatofibroma (DF) in ISIC 2019 dataset.

ISIC 2019 challenge dataset was chosen as the primary dataset instead of the latest ISIC 2020 challenge dataset. This was because the ISIC 2020 dataset consisted only of 7 seven classes instead of the 8 classes of ISIC 2019 and the 2020 dataset focuses primarily on benign and melanoma lesion types which would have limited the scope of this research [31]. Moreover, ISIC 2019 was also a much explored and widely used dataset which provided us with several additional resources to base our research on [32]. Instead of using the dataset directly, a different approach was taken by first segmenting the dataset into classes and then improving upon it by using oversampling and undersampling according to the class size and then applying additional augmentation techniques in the classes where the sample quantity was lower which would prevent the issue of the model being trained with an extra bias for the larger classes. A robust dataset is required for creating an optimal model. So image augmentation was utilized for smaller classes such as AK, DF, VASC, and SCC and reduced the images of larger classes such as NV to match other classes [33]. This would provide a more balanced and accurate result for our model. More of which is discussed in the later part of this methodology section.

2.2. Image Augmentation and Pre-Processing Techniques

Adding to the highly selective and quintessential process of dataset selection, the size of the training dataset also ends up being a determining factor which tends to cause an impact on the model's performance [34]. To counter this problem, the concept of Image Augmentation has been introduced within the steps of Image Pre-Processing, which literally means "to add to images".

2.2.1. Taxonomy

Augmentation can broadly be classified into the following kinds [35]:- (1) Model Free Augmentation: this type of Image Augmentation does not involve the use of a ML/DL model. It includes the use of various Single Image (such as geometrical and intensity transformation, color image processing) and multiple image methods of which the ones relevant to the scope of this paper have been highlighted moving forward [36] :- i) Geometrical Transformation: being one of the simplest ones to execute, this technique still proves to be one of the most effective ones despite its toll on the training process of the model [37]. It includes methods such as translation, rotation, scaling and addition of extra noise. ii) Color Image Processing: most common color image formats (like the RGB and SRGB format) are made up of the combinations of three colors and these colors can be manipulated and tweaked by some randomized values thus forming new augmented images. It includes the use of methods such as color shifting, hue shifting, color inversion and saturation shifting. iii) Intensity Transformation: intensity Transformation techniques such as histogram equalization have also been growing in popularity because of their nature to modify the given image one pixel at a time. This results in an image with varied intensity values which also serves as a viable alternative for image augmentation.

Despite the core advantages of using this specific approach for image augmentation being its simplicity, performance and reliability, one common disadvantage faced due to the use of Model Free Augmentation on a whole is the lack of originality in the newly generated images. This problem is solved by the use of more complex methods of Model Augmentation which bring with themselves their own set of pros and cons.

2.2.2. Model Augmentation

This type of Image Augmentation method involves the use of ML/DL models. Though it is more complicated and requires more resources than other Model Free Augmentation methods, it yields better augmented images with higher originality which is beneficial for the model in most cases [38]. The viable and most common approaches to Model Augmentation relevant to the scope of this paper have been discussed moving forward: i) Generative Adversarial Networks (GANs): the Generative Adversarial Networks (GANs) are composed of a generator and a discriminator in which the sole purpose of the generator is to produce new images and for the discriminator, to classify images and discard the ones that seem "fake". When trained for a sufficient number of "epochs", the network yields results which hold great research value for medical image augmentation [39]. ii) Neural Style Transfer (NST): the Neural Style Transfer (NST) technique utilizes a deep learning model to extract the required features from the image and then simply adding them on another background [40]. This technique is also used for image augmentation which like other methods yields images that have better originality than other Model Free Augmentation Methods.

Although Model Augmentation approaches provide better originality of augmented images by the use of Deep Learning Networks, the problem of huge required resources ends up limiting their use in most studies. Even though they yield images with better originality, they need to be pre-trained and fine tuned in a way particular to an individual research which makes the process time taking as well as complex since without a high accuracy, even simpler approaches out-perform them many a times [41]. A comparatively newer approach is a Optimizing Policy Based Augmentation, which utilizes reinforcement learning and ML/DL neural networks to determine the best strategy for enhancing images [35].

2.2.2. Implementation

Out of the various methods and techniques discussed earlier, the choice of the correct combinations that could help us maximize the productivity of the model and the research was particularly challenging. Considering the limited amount of resources that were available to implement them, some of the complex methodologies including augmentation and preprocessing techniques including the use of ML/DL models stood ruled out. In summary, the techniques employed for Image Augmentation have been summarized in Table 1.

Technique	**Description**	**Purpose**
Translation	Shift the image along the x and y axis randomly.	Helps the model learn the lesion pattern found at different positions within the image.
Scaling	Resizes and crops the image either by zooming in or out by a randomized value.	Helps the model learn the lesion pattern found at different sizes within the image
Rotation	Rotates the image by a randomized angle.	Helps the model learn the lesion pattern variations found in the image by random angles.
Flipping	Creates a mirror image of the original image by flipping horizontally, vertically or both.	Helps the model learn the lesion pattern found at different positions within the image.
Noise Addition	Adds a randomized value of noise in the image using gaussian noise.	Helps model learn lesion patterns even under noise conditions possible in the image due to various factors.

Table 1. *Image Augmentation techniques used in previous research.*

2.2.3. Image Pre-Processing

Preprocessing the acquired image data has shown to significantly improve the accuracy of any image classification model [42]. This involves some standard methods like resizing the image to standard dimensions, normalization of pixel values i.e. bringing all the pixel values between 0-1 which were earlier between 0-255. Then removal of noise, faint hair and unnecessary texturing from the image are also necessary. In terms of dermatological datasets, the images also involve background skin, hair growth over the lesion, etc. which can cause hindrance in image segmentation. Image preprocessing techniques using filters over the images offer a solution to this. For removal of hair from the skin segment gaussian filter, median filters have shown effective results [43][44]. Both Gaussian and median filters were tested, the median filter was able to retain some texture of the Region of Interest while removing the hair properly has lower complexity and is thus employed. To improve the contrast and highlighting the anatomical features in biomedical data, Histogram equalization algorithm is applied over the image to which helps in enhancing the contrast of the image by first plotting a histogram of pixel intensities and then applying cumulative distribution function on the values. This results in an image with higher contrast and reduced noise. Research explored different Histogram Equalisation variants to improve the Medical image classification with CNN and found that Contrast Limited Adaptive Histogram equalization (CLAHE) performed the best and improved contrast while retaining details of the original image [45][46]. Research enhanced images for skin lesion classification using a CLAHE and got a classification rate of above 92% [47]. With the ISIC dataset, CLAHE with a clipLimit of 3.8 and a tile size of (6, 6) gave amazing results, improving contrast between the ROI and the background skin besides, it also helped in improving the segmentation results by preserving the edge information in the image.

Techniques	Description	Purpose
Resizing	The images available in the datasets are of different sizes and resolutions which are converted into (224X224X3)	The model needs a specified size output.
Normalisation	The pixel values are converted in the range of 0-1	Values between 0-1 make it easier for computer to read and process the data
Noise and Hair removal	A median filter with a kernel size 9 is used for this.	Removal of hair from the lesion area is important for proper identification
Histogram Equalization	It increases the contrast based on different pixel intensities in the image.	This improves highlighting lesion areas and edge detection for segmentation of Region of Interest

Table 2. *Image Pre-processing techniques used in previous research.*

2.3. Algorithm for image segmentation and feature extraction

2.3.1. Image Segmentation

The performance of any model directly depends upon the kind of data-fed to it. An important aspect of using deep learning for medical image classification are the robust image segmentation models that have shown to improve the accuracy of medical diagnosis. Image segmentation can be defined as dividing an image into several disjointed areas according to the features like color, texture, shapes, etc. of the image region [48]. The study shows that classification of lung images after segmentation techniques improves model accuracy [49] which can also be seen in skin lesion classification [50]. Another research showed that in medical images, blurriness, segmentation showed a wide range of outcomes with improved accuracy [51].

Two segmentation methods were compared for skin lesions from the given images. (1) U-Net: in 2015, a CNN based model called U-net was introduced for Bio-medical image segmentation. This state-of-the-art model has been used and modified in several researches, involving skin lesion segmentation [50]. Research proposes a highly modified and advanced architecture for a U-net model to segment the region of interest from the background skin where the affected skin area was highlighted with 94% accuracy. Another Research shows the U-net model achieved a jaccard index of 0.80 with the ISIC dataset for segmenting and creating an image mask [52], besides this it also showed that a classification model trained using this segmented data gave more accurate results. We used a model based on U-net with 18 convolutions, 4 Max pooling with Relu activation function, divided into an contracting and expansive path and finally an output layer with sigmoid activation, which converts the input images into a segmented mask. The model was pretrained on a non-medical dataset and then trained on the dermatological dataset PH2 which has lesion segmentation as its groundtruth, this improved the accuracy. Despite having a good accuracy, the model was highly complex and proved to be computationally costly. (2) Thresh Binary: a simpler way of segmentation thresh-binary was explored where the threshold value was set to be 127, the image segments with pixel value greater than that were segmented, treated as white while the background image was assigned the pixel value zero. Color and texture of the affected area are important for skin disease classification thus the segmented area i.e. the pixels with white value are restored from the original image, thus retaining all the essential features of the ROI and removing the background skin. Research shows that segmenting in the surrounding skin is a solution for skin image variability due to factors like skin tone, type, and lighting [53]. Segmenting and preserving the texture of the original lesion image also helped in feature extraction in the image like edge detection, spatial distribution, which are otherwise done by filters like gabor filters [44].

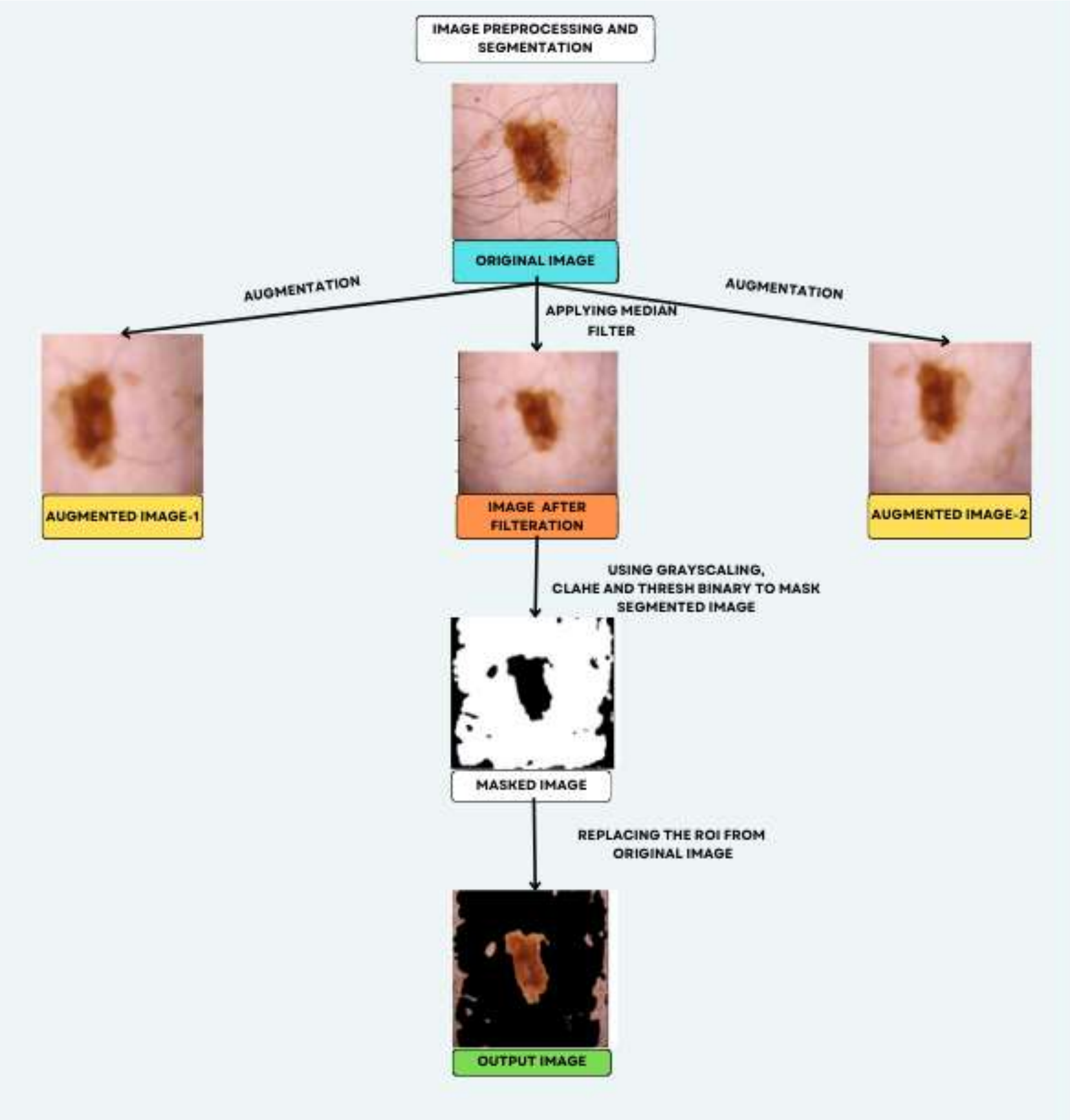

Figure 1. *Presents the proposed image preprocessing and segmentation techniques and their outputs with a test image. The image goes through several rounds of filtering, these filters extract useful features or patterns in order to distinguish between different regions. Due to filtration of image hairs are removed from image. After filtering, a segmentation mask is made. This mask separates the foreground (Skin Lesion) from the background (Surrounding tissue).The final output is an image that only contains segmented objects.*

2.4. Modifying and Assembling Available Models

A plethora of approaches have been explored in identification of skin lesions using unique, robust models and adopting a variety of techniques [54][55].

2.4.1. Transfer Learning

Transfer learning, initially proposed for dealing with lack of robust datasets has been seen as a fascinating approach in skin disease diagnosis, this involves loading a pre trained classifier trained on huge non-medical datasets like imagenets, and then using those as a base model and adding own classification layers, fine tuning the model and re-training on the dataset. These pre-trained models are able to extract features from the new dataset and also reduce computational complexity. Research took a transfer learning approach using the state of the art AlexNet model, modifying it with a softmax classification layer and fine tuning the final model for skin disease detection and were able to achieve an average accuracy of 95% with ISIC dataset [56]. Researchers explored different ResNet models for multiple common dermatological disease classification and were able to achieve a peak accuracy of 89.8% with the ResNet152 model [57]. Research used VGG16 and VGG19 models as transfer learning models and experimented with custom layers, optimizing, activation functions, etc. to identify skin cancer and were able to achieve a Sensitivity of 98.08% with the final proposed model with VGG19 [58].

2.4.2. Two Path Algorithm

Instead of a simple model a two path algorithm has become a topic of research recently. These models are initialized with two different paths taking separate inputs and having separate hidden layers(Convolutions, Pooling, etc.) Their output is combined during the flattening stage which is then given to Fully connected layers for classification [59]. For the purpose of skin disease classification this can be used to advantage, instead of giving a single input image to the model, the preprocessed image along with a segmented ROI image can be given with two different paths and a single classification result can be predicted with higher accuracy.

2.4.3. Ensemble Deep Learning

Ensemble learning is a powerful technique that is a machine learning paradigm which employs multiple machine learning algorithms to train several models [60]. Similarly, Ensemble deep learning is a technique to combine multiple deep learning algorithms to train different models [61]. This process of combining is also known as stacking in which the new model learns how to best combine the predictions from multiple models [62]. The advantage of this is that it fuses the result from these models and utilizes voting schemes to achieve knowledge discovery and better predictive performance than any individual model can produce along [63]. This study uses ensemble learning to combine the strengths of state of the art models like - EfficientNetB2V2, MobileNetV2 and InceptionV3 to obtain a better result. Two models were required for a proposed two path approach so transfer learning using different models was inspected to find out which two models worked best together. This stacking allowed for an improved predictive performance, increased robustness to errors and uncertainties, and reduced overfitting by combining the results [64][65].

2.4.4. Optimizers

Optimizers are algorithms used to find the optimal set of parameters set for a model during the training process. Stochastic Gradient Descent (SGD) is a powerful optimization algorithm in which the actual gradient is replaced by an estimate calculated from a randomly selected subset of a data [66]. Another one of the optimizers is Adam Optimizer which is a popular optimization algorithm used in deep learning for training neural networks. While they share some similarities, they differ in their learning rate, momentum, and adaptability to different learning rates. The SGD is shown to give better generalization but does not handle bad hyperparameters like learning rate well. Research compares the two optimizers with VGG16 and VGG19 models for skin cancer diagnosis, the results showed that the Adam optimizer gave significantly better outputs while maintaining constant accuracy [58]. For the scope of the research, we used Adam optimizer as it is easy to implement, has an adaptive learning rate, faster convergence, and enhanced performance for our model [67][68].

2.4.5. Learning Rate

Learning rate in deep learning models is a critical hyperparameter which governs the pace at which the model learns and updates its weights during the training process.
The paper has used a low learning rate to ensure that it provides a stable solution, avoiding divergence, and improved generalization accuracy for our model [69].

2.5. Implemented Model Architecture

Separate image preprocessing and segmentation and preprocessing methods were used to create two images from a given image and for the two path model to work the ImageDataGenerator() function is used to first load the raw image data along with the class label. Then each goes through both methods - preprocessing and segmentation, returning two images. Then the DataGenerator function finally yields Tensorspec objects inside a tuple where the preprocessed image with its corresponding segmentation mask is stored with their single label. This tuple is passed into the tf.data.Dataset API to return the final dataset which would provide two input images to the model.

EfficientNetB2V2 and MobileNetV2 models with their pre-trained weights for the ImageNet data were used as two base models for base feature extraction because both the models can work with very low computational resources and the input and output shapes of the original models are similar. Inception V3 despite giving promising results was discarded as the other models were more compatible.

The EfficientNetV2S model has been used for the first path with simply preprocessed images. The model uses pointwise as well as depth wise convolutions with ReLu activation and batch normalization in an inverted bottleneck block architecture. The model comprises several such blocks with varying filters, kernel size according to the requirement. The last fully connected classification layers have been removed, then all the other layers of the base model are frozen so that the pre-trained weights are not optimized in the first cycles. A convolution layer with 3 by 3 kernel and 128 filters and a 2 by 2 max pooling layer is added to the base model output. This output is flattened so that it can be given to the ensemble model for classification by using fully connected dense layers [70].

MobileNetV2 is an efficient neural network that offers improved performance and scalability for mobile and embedded devices. It uses a novel layer called an inverted residual block; unlike traditional residual blocks, input and output channels are expanded in the inverted residual block while the intermediate layers have a smaller number of channels. It utilizes a linear activation function in the bottleneck layers of the inverted residual blocks. It also splits the convolution operation into depthwise convolution and a pointwise convolution which reduces the computational costs [71]. Moreover, this model is designed to be efficient in terms of both memory usage and computational costs.

After flattening, both models return a one dimensional array output which is suitable for the fully connected classification layer.

For the final ensembling model, the outputs from the two paths i.e. the MobileNetV2 and EfficientNetB2V2 are concatenated. A fully connected layer with 1024 output dimension using a softmax activation function is added to pass the concatenated features. A dropout layer with 0.5 rate is applied to reduce overfitting, before declaring the final output layer with softmax activation which gives output in form of a probability from the 8 given classes. A softmax activation function is typically used in the final classification layer; it converts the output into probabilities for different classes, thus yielding better results for classification models.

This model is trained for 10 epochs with frozen layers which keep the weights of the base models same, only changing the weights of added layers, then for fine tuning the base models with our data, the layers of both the base models are unfrozen and the model is trained again for 10 epochs with a low learning rate so as to maintain the classification performance of the base models.

2.6. Performance and Evaluation Indexes

The classification performance of our model was evaluated by measuring the following metrics: Accuracy, Precision, Recall, F1 Score and Categorical Cross-Entropy Loss. The description of the used metrics are as follows.

2.6.1. Accuracy

Accuracy is defined as the proportion of correctly predicted samples and the total number of sample predictions.

$$[\ \text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP} \times 100\ ]$$

2.6.2. Precision

This refers to the proportion of correctly predicted positive samples and the total number of positive sample predictions.

$$[\ \text{Precision} = \frac{TP}{TP+FP} \times 100\ ]$$

2.6.3. Recall

The recall, also known as sensitivity or TPR, is defined as the proportion of correctly predicted positive samples out of the total number of actual positive samples. It measures the ability of the model to identify all relevant instances within the dataset.

$$[\ \text{Recall} = \frac{TP}{TP+FN} \times 100\ ]$$

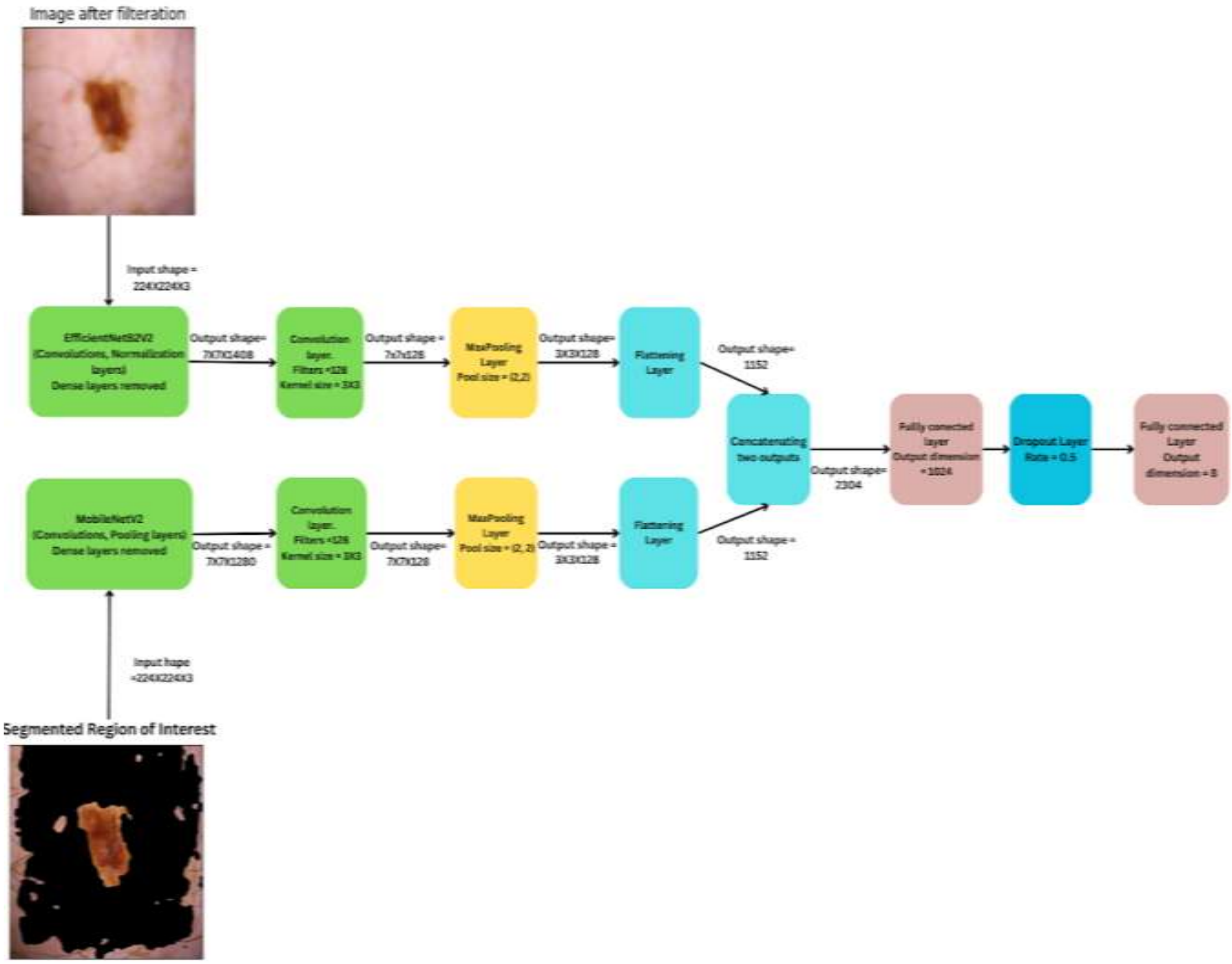

Figure 2. *Layer architecture of the final CNN model proposed, highlighting the two base transfer learning models which act as feature extractors for the same image preprocessed in two different ways as depicted. The two path architecture is displayed with the two models and additional convolutions, max pooling layer. In both the models dense layers are removed from the original model. As shown, the output from both the paths is flattened and concatenated which creates a new 2304 dimensional tensor. Two fully connected layers are used to identify complex patterns within extracted features as the first layer has 1024 units along with a dropout layer to reduce overfitting and the final output layer with softmax activation gives a prediction from the 8 classes*

2.6.4. F1 Score

The F1 Score refers to the weighted average of the precision and the recall, combined into a single value to understand the performance of the model better.

[F1 Score = $2\frac{Precision \times Recall}{Precision + Recall} \times 100$]

2.6.5. Categorical-Cross Entropy Loss

This metric measures the performance of a classification model, where the prediction is a probability value between 0 to 1. It is a loss function used in multi-class classification problems, where a model assigns an image to only one of several possible classes, and is preferred due to its simplicity, effectiveness and ease of implementation. Generally, the lesser the log loss, the better the model performs.

[L = $\frac{-1}{N}\sum_{i=1}^{N} \sum_{c=1}^{C} yi,jlog(Pi,j)$]f

Where:

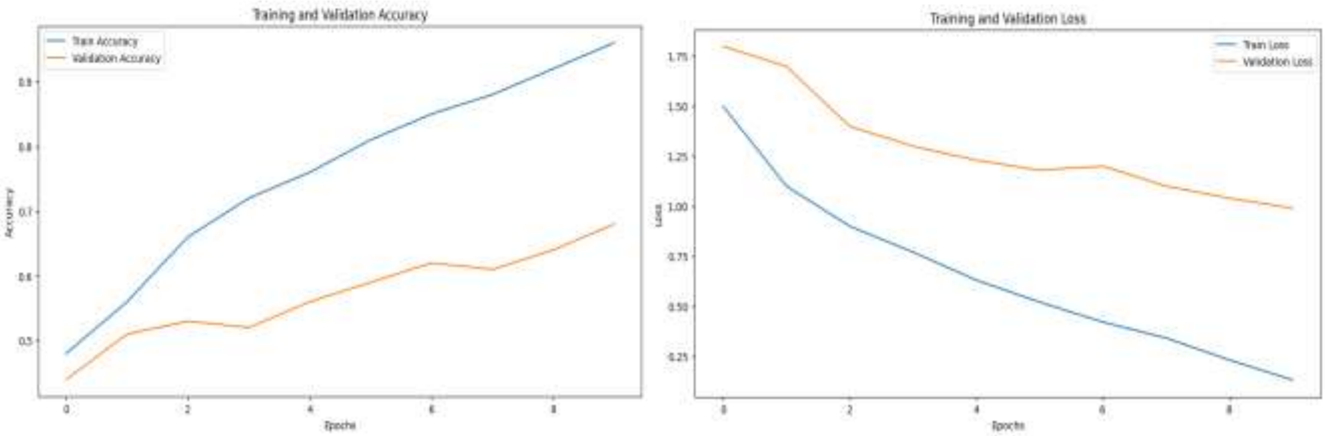

Figure 3. *(a) Accuracy vs No. of Epochs graph which gives a clear overview of the increase in train and val accuracy and a comparison between the two during the model training. (b) Loss vs No. of Epochs graph that depicts the comparison between train and val categorical cross entropy loss during the model training.*

L is the categorical-cross entropy loss
N is the number of samples
C is the number of classes
$\mathbf{Y}_{\mathbf{i,J}}$ is a true label for the i^{th} sample and j^{th} class
$\mathbf{P}_{\mathbf{i,J}}$ is a predicted probability for the i^{th} sample and j^{th} class

2.7. Machine-Learning-Based Skin Disease Detection and Classification

For the scope of this research, the experiment was done on a python platform 3.10.14 and the entire model was coded on tensorflow library using the keras API. The primary dataset utilized during this experiment was the ISIC 2019 Challenge Dataset containing 25,531 images of eight different classes.
Combining all the discussed techniques, the paper has compiled an ensemble learning model which uses a two path algorithm utilizing already existing models: EfficientNetB2V2 and MobileNetV2 which are computationally simpler than other models used for this purpose previously. To get best weights for the proposed model experimentation was done with different optimizers, number of epochs, early stopping and fine tuning of the transfer learning models. Final weights for the model were obtained after running it for 10 epochs. To make sure that the model does not overfit and saves time during training, a callback option was placed for early stopping to analyze and save the weights after epochs which gave the best validation accuracy. Instead of using a separate fine tuning compilation, some of the layers from both the models were kept frozen to maintain while the last 30 layers were unfrozen, this improved the train accuracy significantly which was earlier just 75%. For testing the effectiveness and real usability of the model, a test data was created containing images from different classes for the model to predict to measure the model's performance based on different standard metrics.

3. Results

After running the model for 10 epochs, a train accuracy of 95.23% was reached along with a train loss of 0.1367 from an initial train accuracy of 38.02% along with an initial train loss of 1.572 in the first epoch. The validation accuracy attained on the other hand was 68% and a validation loss of 0.99 (figure 3). After testing the model using the most accurate weights with the images from different classes, the test accuracy of the model was found to be about 0.70 . The recall or sensitivity of the model was calculated out to be 0.75 and the precision of the model was found out to be 0.625. This gave an overall F1 score of approximately 0.667 for the model. The confusion matrix in figure 4 shows the model performance with distinct features while two classes that seem to confuse the model are AK (Actinic Keratosis) and DF (Dermatofibroma). On further exploring the data in these classes, it was found that the images lacked distinctly recognizable features and had several distinct independent variations of the disease which seemed to be a bottleneck for the model's learning. The images in the classes such as VASC with a limited amount of data were augmented and this has proved to yield results as high as 93%. Using EfficientNet and MobileNet as base feature extractors, the model gives promising results with lower computational complexity compared to the other state of the art models. The proposed model outperforms many models in terms of the training accuracy with the ISIC-2019 dataset and can be taken for further testing with more data for more significant results.

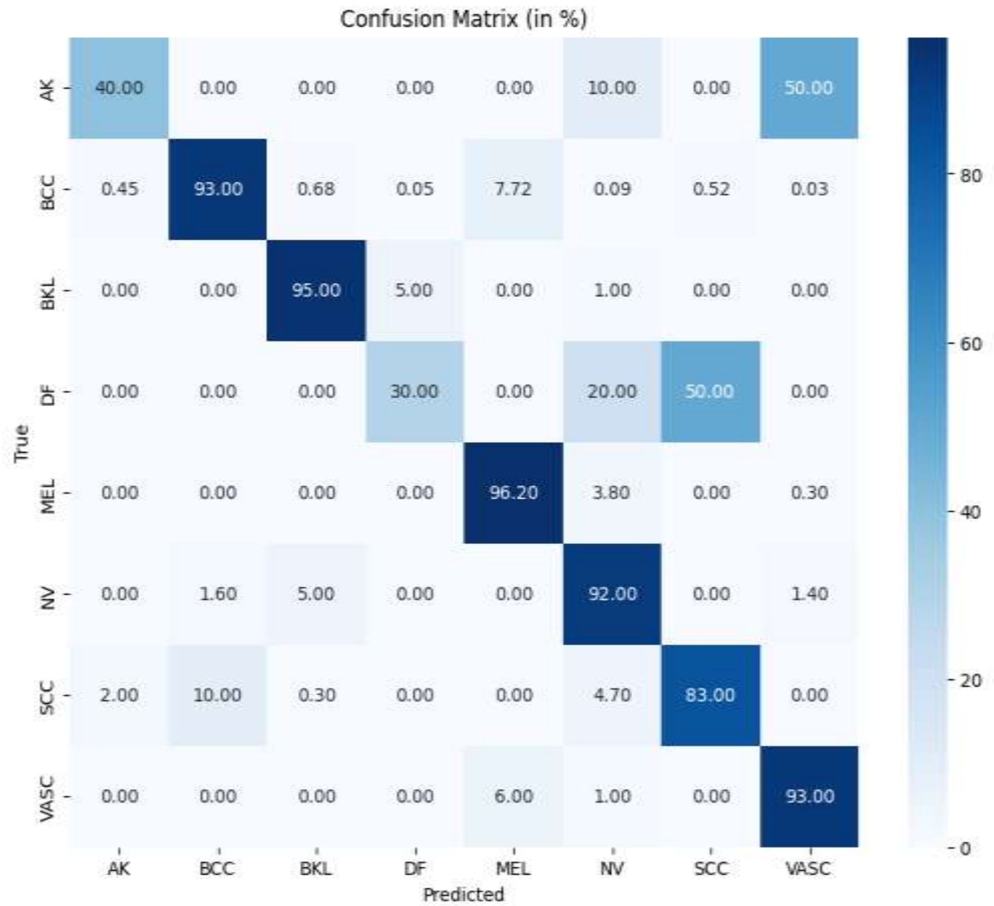

Figure 4. *Confusion matrix by giving the model equal number of images from each of the eight classes.*

Using a two-path algorithm for separate processed image and segmented image feature extraction helps deal with several issues that arise due to variation in images in skin disease classification, using a normal pre processed image along with a segmented ROI image helps the model generalize and concatenating the two outputs before classification helps in dealing with overfitting problem. Using the segmented ROI also tackles the problem of different skin-tones which normally cause a hindrance for the model and make it difficult to use in the real world diagnosis.

The ISIC 2019 dataset is one of the most robust dataset for multiple-disease class identification in dermatology but the dataset is far from perfect. Although the dataset offers high quality of images, the classes are highly imbalanced where half the images in the dataset belong to a single class while the smallest class only has 200 images. Despite using image augmentation in an attempt to mitigate the issue, the confusion matrix clearly shows that some classes are better identified than others. Exploring further into the dataset it was found that few classes in the dataset contain images which can obstruct the feature extracting models during training and validation because the lesion areas in these images are either completely covered or very minute as compared to the other images in the same class. It can also be observed that dermatological classification with more than 3-5 classes from the ISIC dataset gives significantly lower results as most of the features like inflammation, rough and dried texture, scaly skin, etc are extremely similar in many classes. This causes confusion in the feature extractors and leads to inaccurate results when predicting labels for new images. Thus, the research relies on modification and addition to the base dataset from other available datasets.

3.1. Challenges and Limitations

Unlike a binary classification problem, solving the multi-class dermatological diseases classification problem has several complexities. The foremost problem which arises is the availability of required images for each disease. Currently available datasets for multi class dermatological

Research Paper	Deep Learning Method	Models	Dataset	Target Classes	Accuracy of the Best Model
Z Rahman et. al., 2021 [72]	Ensemble learning with weighted averaging.	ResNeXt, SeResNeXt, ResNet, Xception, and DenseNet.	HAM10000 ISIC-2019	AKIEC, BCC, BKL, DF, MEL, NV, VASC,	88.0%
Kemal Polat et. al., 2020 [25]	One-Versus-All approach & Ensemble learning	Deep CNN	HAM10000	AKIEC, BCC, BKL, DF, MEL, NV, VASC,	92.9%
AR. Lopez et. al.,2017 [73]	Transfer learning	VGG 16	ISIC Archive	Malignant vs benign skin lesions.	81.3%
Catarina Barata et. al., 2019 [74]	Hierarchical Attention mechanism(CNN-LSTM) and Ensembling	DenseNet-161 and ResNet-Inception	ISIC-2017 ISIC-2018	Melanocytic vs Non-Melanocytic MEL, NV, BCC, AKIEC, BKL, DF, and VASC.	81.3%(balanced multi-class accuracy) 64.1%(balanced multi-class accuracy)
Nils Gessert et. al., 2018 [75]	Transfer Learning, Ensembling and Meta-Learning	54 models with these architectures: SENet154, ResNeXt101 32x4d, Densenet201, Densenet161, Densenet169, SE-Resnet101, PolyNet.	ISIC-2016, ISIC-2017 and HAM10000.	MEL, NV, BCC, AKIEC, BKL, DF, and VASC.	97.0% 85.6%(balanced multi-class accuracy)
Amirreza Mahbod et. al., 2020 [76]	Transfer Learning, and Ensembling(Multi-Scale Multi CNN 3 level fusion Scheme)	EfficientNetB0, EfficientNetB1 and SeResNeXt-50	ISIC-2016, ISIC-2017, and ISIC-2018,	MEL, NV, BCC, AKIEC, BKL, DF, and VASC.	96.1% 87.4%(balanced multi-class accuracy)
Sara Atito Ali Ahmed1 et. al., 2020 [77]	Transfer Learning, Model Fine-tuning and Ensembling	Xception, Inception-ResNet-V2, and NasNetLarge	ISIC-2019	MEL, NV, BCC, AK, BKL, DF, VASC, SCC, and UNK(None of the others)	91.5% 60.2%(balanced multi-class accuracy)
Josef Steppan et. al., 2021 [78]	Transfer Learning, Model Fine Tuning and Ensembling	EfficientNet-B5, SE-ResNeXt-101(32x4d),	ISIC-2019, PH2, 7 Point criteria	MEL, NV, BCC, AK, BKL, DF, VASC, SCC, and	92.3% 63.4%(balanced multi-class

		EfficientNet-B4, Inception-ResNet-v2	database, MED-NOD, SKINL2 and SD-198.	UNK(None of the others)	accuracy)
Esteva et al., 2017 [23]	Transfer Learning	Inception v3	ISIC Archive, DermNet, SD-198	Melanoma, BCC, SCC	72.1%
Our model presented in this research	Transfer Learning and Ensembling	EfficientNetB2V2, MobileNetV2	ISIC-2019 + Additional data	AK, BCC, DF, BKL, MEL, NV, SCC, VASC	70.3%

Table 3. *Past image classification techniques used and results.*

disease classification like the ISIC, HAM1000, DermNet, etc. all are based on specific classes of diseases and contain dermoscopic images. Due to this the real world accuracy of the model remains low and the model cannot precisely diagnose the specific disease in the class. Although modifying these datasets and building a dataset with more classes can help in training the model to classify more diseases, increasing the number of classes can negatively affect the model's accuracy.

Another issue that arises in dermatoscopic classification is presence of multiple lesion areas in a single image, which is commonly seen in the real world. Segmenting the ROI has been proposed as a solution for this but in most cases the less significant lesion goes undetected. Besides the quality of the images, it can also be stated that the different angles and lighting conditions affect the practical usage of models with new data. The proposed solution to this is creating variations of the same image using augmentation techniques which rotate the images by random angles, randomly flip the image, etc. The field of view of the dermoscope can affect the visibility of the lesion in the image. This can adversely affect the classification accuracy, the proposed solution to this in the research involves modification of the dataset to have a variation of images with a variety of FOVs and difference in the overall spread of the lesion area. Many skin diseases are found in the scalp region or are otherwise covered with hair. Although the simpler approach as proposed in the paper i.e. using filters over the image during preprocessing give promising results for small, thin patches of hair over the lesion but these can tamper with and even remove the entire lesion area when the ROI is densely covered with hair. It is also crucial to point out that several skin diseases cannot be identified just based on the images due to similarity between different diseases, a reliable diagnosis requires background details and other tests which an image classification model ignores. These models work under the assumption that the different classes have different features for identification but that is not the case with several skin diseases. This can lead to false diagnosis that can be dangerous in the real world.

4. Conclusion

Analyzing the problems, the different solutions given by researchers in the field and their limitations a future pathway is formed to completely overcome the complex skin disease classification problem. Following are the most prominent and promising segments in this field that can lead to a bio-technical breakthrough. Several models have been able to achieve amazing accuracy for classification on a few classes such as those available in the ISIC datasets but the real skin disease diagnosis problem is far from being solved because all these models assume the availability of high quality dermoscopic images which do not go beyond the specified class and have exorbitant computational cost. For a real world classifier it is important to ensure that more common diseases can be diagnosed with high accuracy using simple camera images. For this the SCIN dataset seems very promising where all the data has been collected through online ads from common people.

A classifier that considers a patient's medical records and other factors like age, gender, etc. can be a step into the future of dermatological disease classification. A single dermatological disease class can have several dissimilar types of images with different shapes, textures and color patterns images. When the features for a single class are so different, it makes multi class identification almost with high accuracy almost impossible. An unexplored approach for this is subdividing similar images inside a class into subclasses, such that the model treats them as different classes with the same label name. This approach can be taken for data acquisition and processing for training a model that diagnoses a large number of skin diseases with easily available, low quality images. The most prominent future direction is moving towards an accessible model which can identify most of the common and dangerous diseases using the insights shared in the research.

Deep Learning, a more advanced and modern approach to Machine Learning has produced more efficient and reliable results over the predecessor almost every time [79]. The classification of medical images into the ones affected by various dermatological diseases via the use of such deep learning models has been a recent topic of research but most of these models provide binary based classification, focusing only on identifying if the patient is infected with a specific disease only [53]. This has been a hindrance to the reliability of artificial intelligence models for the dermatological diagnosis problem. The ISIC-2019 Challenge Dataset used in the research offers an eight class classification of dermatological diseases but a prominent problem of class imbalance is still persistent. Furthermore, the images of some classes contain images that often contribute to bringing down the model's accuracy because of their dissimilar nature and lack of consistent features. A proposed solution for the same is grouping the dataset classes into further subclasses using unsupervised learning [80]. That could help provide a more specific diagnosis. The power of State-of-the-art models such as InceptionV3, ResNet and VGG19 can also be harnessed for more reliable and accurate results. A major aspect of this study is the use of ensemble learning along with a two path algorithm, which has helped boost the performance of the proposed model and can also be further explored with other high performing models to achieve notable results. The model's performance could further benefit from a deeper dive into better image preprocessing and feature extraction methods [81]. This can help produce results comparable to other complex models even on lower-end devices further assisting the ease of early detection of these diseases by medical professionals in the fields of single cell genomics, biosensors as well as non-medical professionals [82] [83] [84].

Author Contributions

R.C: Software, Methodology, Writing, Validation. A.B: Investigation, Data curation, Writing, Methodology, Software. A.K: Investigation, Data curation, Writing, Methodology. S.H: Formal Analysis, Writing, Visualisation. G.B: Formal Analysis, Visualization, Writing. A.S, K.V., K.D., S.V. - conceptualization, original draft- reviewing and editing, final draft- reviewing and editing.

Conflicts of Interest

The authors declare no competing financial interests or conflicts of interest.

References

[1] National Cancer Institute, "Skin cancer (including melanoma)—patient version," *National Cancer Institute*, 2019. https://www.cancer.gov/types/skin

[2] M. Arnold *et al.*, "Global Burden of Cutaneous Melanoma in 2020 and Projections to 2040," *JAMA Dermatology*, vol. 158, no. 5, pp. 495–503, Mar. 2022, doi: https://doi.org/10.1001/jamadermatol.2022.0160.

[3] "Skin Cancer Facts & Statistics," *The Skin Cancer Foundation*. https://www.skincancer.org/skin-cancer-information/skin-cancer-facts/#general

[4] Q. Liu, M. Sha, B. Xue, L. Shen, G. Li, and X. Cheng, "Health-related quality of life and associated factors among non-melanoma skin cancer patients: a cross-sectional study.," vol. 11, no. 3, pp. 150–150, Feb. 2023, doi: https://doi.org/10.21037/atm-22-6654.

[5] S.-Y. Z. Kao, D. U. Ekwueme, D. M. Holman, S. H. Rim, C. C. Thomas, and M. Saraiya, "Economic burden of skin cancer treatment in the USA: an analysis of the Medical Expenditure Panel Survey Data, 2012–2018," *Cancer Causes & Control*, Nov. 2022, doi: https://doi.org/10.1007/s10552-022-01644-0.

[6] CDC, "Health and Economic Benefits of Skin Cancer Interventions," *National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP)*, May 24, 2024. https://www.cdc.gov/nccdphp/priorities/skin-cancer.html

[7] H. Kittler, H. Pehamberger, K. Wolff, and M. Binder, "Diagnostic accuracy of dermoscopy," *The Lancet. Oncology*, vol. 3, no. 3, pp. 159–65, 2002, doi: https://doi.org/10.1016/s1470-2045(02)00679-4.

[8] H. J. Pandya et al., "Label-free electrical sensing of bacteria in eye wash samples: A step towards point-of-care detection of pathogens in patients with infectious keratitis," Biosensors and Bioelectronics, vol. 91, pp. 32–39, May 2017, doi: https://doi.org/10.1016/j.bios.2016.12.035.

[9] S. Odinotski et al., "A Conductive Hydrogel-Based Microneedle Platform for Real-Time pH Measurement in Live Animals," Small, vol. 18, no. 45, Sep. 2022, doi: https://doi.org/10.1002/smll.202200201.

[10] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big Data*, vol. 6, no. 1, Dec. 2019, doi: https://doi.org/10.1186/s40537-019-0276-2.

[11] M. Rana and M. Bhushan, "Machine learning and deep learning approach for medical image analysis: diagnosis to detection," *Multimedia Tools and Applications*, vol. 82, Dec. 2022, doi: https://doi.org/10.1007/s11042-022-14305-w.

[12] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, Mar. 2021, doi: https://doi.org/10.1186/s40537-021-00444-8.

[13] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning," *arXiv:1602.07261 [cs]*, Aug. 2016, Available: https://arxiv.org/abs/1602.07261

[14] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv.org*, Apr. 10, 2015. https://arxiv.org/abs/1409.1556

[15] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated Residual Transformations for Deep Neural Networks," *IEEE Xplore*, Jul. 01, 2017. https://ieeexplore.ieee.org/document/8100117/

[16] B. Khasoggi, E. Ermatita, and S. Samsuryadi, "Efficient mobilenet architecture as image recognition on mobile and embedded devices," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 16, no. 1, p. 389, Oct. 2019, doi: https://doi.org/10.11591/ijeecs.v16.i1.pp389-394.

[17] G. Huang, Z. Liu, and Weinberger, Kilian Q, "Densely Connected Convolutional Networks," *arXiv.org*, 2016. https://arxiv.org/abs/1608.06993

[18] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *arXiv.org*, 2019. https://arxiv.org/abs/1905.11946

[19] Md. Z. Islam, Md. M. Islam, and A. Asraf, "A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images," Informatics in Medicine Unlocked, vol. 20, p. 100412, 2020, doi: https://doi.org/10.1016/j.imu.2020.100412.

[20] Aditi, Mayank Kumar Nagda and Poovammal E"Image Classification using a Hybrid LSTM-CNN Deep Neural Network," *International Journal of Engineering and Advanced Technology*, vol. 8, no. 6, pp. 1342–1348, Aug. 2019, doi: https://doi.org/10.35940/ijeat.f8602.088619.

[21] H. Wu *et al.*, "CvT: Introducing Convolutions to Vision Transformers," *arXiv:2103.15808 [cs]*, Mar. 2021, Available: https://arxiv.org/abs/2103.15808

[22] Kulkarni S, Dhingra K, Verma S., "Applications of CMUT Technology in Medical Diagnostics: From Photoa-coustic to Ultrasonic Imaging", International Journal of Science and Research (IJSR), Volume 13 Issue 6, June 2024, pp. 1264-1269, https://www.ijsr.net/ar-chive/v13i6/SR24619062609.pdf.

[23] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Jan. 2017, doi: https://doi.org/10.1038/nature21056.

[24] H. A. Haenssle *et al.*, "Man against machine: diagnostic performance of a deep learning convolut ional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of Oncology*, vol. 29, no. 8, pp. 1836–1842, May 2018, doi: https://doi.org/10.1093/annonc/mdy166.

[25] K. Polat and K. Onur Koc, "Detection of Skin Diseases from Dermoscopy Image Using the combination of Convolutional Neural Network and One-versus-All," *Journal of Artificial Intelligence and Systems*, vol. 2, no. 1, pp. 80–97, 2020, doi: https://doi.org/10.33969/ais.2020.21006.

[26] M. N. Bajwa *et al.*, "Computer-Aided Diagnosis of Skin Diseases Using Deep Neural Networks," *Applied Sciences*, vol. 10, no. 7, p. 2488, Apr. 2020, doi: https://doi.org/10.3390/app10072488.

[27] H. M. Son *et al.*, "AI-based localization and classification of skin disease with erythema," *Scientific Reports*, vol. 11, no. 1, p. 5350, Mar. 2021, doi: https://doi.org/10.1038/s41598-021-84593-z.

[28] M. T, F. Pm, M. Js, M. Ar, and R. J, "PH2 - A Dermoscopic Image Database for Research and Benchmarking," *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, 2013. https://pubmed.ncbi.nlm.nih.gov/24110966/

[29] P. Tschandl, C. Rosendahl, and H. Kittler, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, no. 1, Aug. 2018, doi: https://doi.org/10.1038/sdata.2018.161.

[30] GhavamiNejad P, GhavamiNejad A, Zheng H, Dhingra K, Samarikhalaj M, Poudineh M., "A Conductive Hydrogel Mi-croneedle-Based Assay Integrating PEDOT:PSS and Ag-Pt Nanoparticles for Real-Time, Enzyme-Less, and Electro-chemical Sensing of Glucose," Advanced Healthcare Materials, vol. 12, no. 1, Oct. 2022, doi: https://doi.org/10.1002/adhm.202202362.

[31] V. Rotemberg *et al.*, "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," *Scientific Data*, vol. 8, no. 1, p. 34, Jan. 2021,

doi: https://doi.org/10.1038/s41597-021-00815-z.

[32] B. Cassidy, C. Kendrick, A. Brodzicki, J. Jaworek-Korjakowska, and M. H. Yap, "Analysis of the ISIC image datasets: Usage, benchmarks and recommendations," *Medical Image Analysis*, vol. 75, p. 102305, Jan. 2022, doi: https://doi.org/10.1016/j.media.2021.102305.

[33] Periketi, P.; Kaur, K. .; Naseer Vaid, F. .; Sree M, Y. .; Madhu, M. .; Verma, S. .; Dhingra, K. . Blood Brain Barrier-on-a-Chip Permeation to Model Neurological Diseases Using Microfluidic Biosensors. J. Knowl. Learn. Sci. Technol. 2024, 3 (4), 78-93. https://doi.org/10.60087/jklst.v3.n4.p78.

[34] H. J. Pandya et al., "A microfluidic platform for drug screening in a 3D cancer microenvironment," Biosensors and Bioelectronics, vol. 94, pp. 632–642, Aug. 2017, doi: https://doi.org/10.1016/j.bios.2017.03.054.

[35] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A Comprehensive Survey of Image Augmentation Techniques for Deep Learning," *Pattern Recognition*, vol. 137, p. 109347, May 2023, doi: https://doi.org/10.1016/j.patcog.2023.109347.

[36] M. Safavieh et al., "Paper microchip with a graphene-modified silver nano-composite electrode for electrical sensing of microbial pathogens," Nanoscale, vol. 9, no. 5, pp. 1852–1861, 2017, doi: https://doi.org/10.1039/c6nr06417e.

[37] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. 1, Jul. 2019, doi: https://doi.org/10.1186/s40537-019-0197-0.

[38] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, p. 100258, Nov. 2022, doi: https://doi.org/10.1016/j.array.2022.100258.

[39] Y. Chen *et al.*, "Generative Adversarial Networks in Medical Image augmentation: A review," vol. 144, pp. 105382–105382, Mar. 2022, doi: https://doi.org/10.1016/j.compbiomed.2022.105382.

[40] J. Li, Q. Wang, H. Chen, J. An, and S. Li, "A Review on Neural Style Transfer," *Journal of Physics: Conference Series*, vol. 1651, no. 1, p. 012156, Nov. 2020, doi: https://doi.org/10.1088/1742-6596/1651/1/012156.

[41] L. Nanni, M. Paci, S. Brahnam, and A. Lumini, "Comparison of Different Image Data Augmentation Approaches," *Journal of Imaging*, vol. 7, no. 12, p. 254, Nov. 2021, doi: https://doi.org/10.3390/jimaging7120254.

[42] M. Ahammed, Md. A. Mamun, and M. S. Uddin, "A machine learning approach for skin disease detection and classification using image segmentation," *Healthcare Analytics*, vol. 2, p. 100122, Nov. 2022, doi: https://doi.org/10.1016/j.health.2022.100122.

[43] E. Jana, R. Subban, and S. Saraswathi, "Research on Skin Cancer Cell Detection Using Image Processing," *IEEE Xplore*, Dec. 01, 2017. https://ieeexplore.ieee.org/document/8524554

[44] A. Singh, A. Kumar Sachan, K. Singh, V. Singh, Sunil, and S. Mishra, "Skin Disease Detection using Image Processing," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 5387, pp. 2582–5208, doi: https://doi.org/10.56726/IRJMETS39556.

[45] Shoffan Saifullah and Rafał Dreżewski, "Modified Histogram Equalization for Improved CNN Medical Image Segmentation," *Procedia computer science*, vol. 225, pp. 3021–3030, Jan. 2023, doi: https://doi.org/10.1016/j.procs.2023.10.295.

[46] N. Salem, H. Malik, and A. Shams, "Medical image enhancement based on histogram algorithms," *Procedia Computer Science*, vol. 163, pp. 300–311, 2019, doi: https://doi.org/10.1016/j.procs.2019.12.112.

[47] J. Premaladha and K. S. Ravichandran, "Novel Approaches for Diagnosing Melanoma Skin Lesions Through Supervised and Deep Learning Algorithms," *Journal of Medical Systems*, vol. 40, no. 4, p. 96, Apr. 2016, doi: https://doi.org/10.1007/s10916-016-0460-2.

[48] X. Liu, L. Song, S. Liu, and Y. Zhang, "A Review of Deep-Learning-Based Medical Image Segmentation Methods," *Sustainability*, vol. 13, no. 3, p. 1224, Jan. 2021, doi: https://doi.org/10.3390/su13031224.

[49] S. Gite, A. Mishra, and K. Kotecha, "Enhanced lung image segmentation using deep learning," *Neural Computing and Applications*, Jan. 2022, doi: https://doi.org/10.1007/s00521-021-06719-8.

[50] Gupte, P.; Dhingra, K.; Saloni , V. Precision Gene Editing Strategies With CRISPR-Cas9 for Advancing Cancer Immunotherapy and Alzheimer's Disease. J. Knowl. Learn. Sci. Technol. 2024, 3 (4), 11-21. https://doi.org/10.60087/jklst.v3.n4.p11.

[51] Pradeep Kumar Singh, Y. Singh, M. H. Kolekar, Arpan Kumar Kar, Jitender Kumar Chhabra, and A. Sen, *Recent Innovations in Computing*. Springer Science+Business Media, 2021. doi: https://doi.org/10.1007/978-981-15-8297-4.

[52] Z. A. Nazi and T. A. Abir, "Automatic Skin Lesion Segmentation and Melanoma Detection: Transfer Learning Approach with U-Net and DCNN-SVM," *Proceedings of International Joint Conference on Computational Intelligence*, pp. 371–381, Jul. 2019, doi: https://doi.org/10.1007/978-981-13-7564-4_32.

[53] L. AlSuwaidan, "Deep Learning Based Classification of Dermatological Disorders," *Biomedical Engineering and Computational Biology*, vol. 14, p. 11795972221138470,

Jul. 2023, doi: https://doi.org/10.1177/11795972221138470.

[54] Pundlik, A.; Verma, S.; Dhingra, K. Neural Pathways Involved in Emotional Regulation and Emotional Intelligence. J. Knowl. Learn. Sci. Technol. 2024, 3 (3), 165-192. https://doi.org/10.60087/jklst.vol3.n3.p.165-192.

[55] H. Li, Y. Pan, J. Zhao, and L. Zhang, "Skin disease diagnosis with deep learning: a review," *arXiv.org*, Dec. 06, 2020. https://arxiv.org/abs/2011.05627

[56] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Classification of skin lesions using transfer learning and augmentation with Alex-net," *PLOS ONE*, vol. 14, no. 5, p. e0217293, May 2019, doi: https://doi.org/10.1371/journal.pone.0217293.

[57] S. Mishra, H. Imaizumi, and T. Yamasaki, "Interpreting Fine-Grained Dermatological Classification by Deep Learning," Jun. 2019, doi: https://doi.org/10.1109/cvprw.2019.00331.

[58] A. Faghihi, M. Fathollahi, and R. Rajabi, "Diagnosis of Skin Cancer Using VGG16 and VGG19 Based Transfer Learning Models," *arXiv.org*, Apr. 01, 2024. https://arxiv.org/abs/2404.01160

[59] D. S. Charan, H. Nadipineni, S. Sahayam, and U. Jayaraman, "Method to Classify Skin Lesions using Dermoscopic images," *arXiv.org*, Dec. 20, 2022. https://arxiv.org/abs/2008.09418.

[60] B. Naderalvojoud and T. Hernandez-Boussard, "Improving machine learning with ensemble learning on observational healthcare data," *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2023, pp. 521–529, 2024, Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10785929/

[61] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," *Engineering Applications of Artificial Intelligence*, vol. 115, p. 105151, Oct. 2022, doi: https://doi.org/10.1016/j.engappai.2022.105151.

[62] W. Wan, Q. Xu, H. Chen, and Q. Chen, "Using Deep Learning Neural Networks and Stacking Ensemble Learning to Predict CSI 300 Index," *IEEE Xplore*, Oct. 01, 2022. https://ieeexplore.ieee.org/document/9978482.

[63] Dong, Xibin, et al. "A Survey on Ensemble Learning." *Frontiers of Computer Science*, vol. 14, no. 2, 30 Aug. 2019, pp. 241–258, https://doi.org/10.1007/s11704-019-8208-z.

[64] Ashok, L R, et al. "Detection of Macular Diseases from Optical Coherence Tomography Images: Ensemble Learning Approach Using VGG-16 and Inception-V3." Algorithms for Intelligent Systems, 1 Jan. 2021, pp. 101–116, https://doi.org/10.1007/978-981-16-3246-4_9. Accessed 28 June 2024.

[65] Sowmiya Baskar, et al. "An Ensemble of Transfer Learning Based InceptionV3 and VGG16 Models for Paddy Leaf Disease Classification." ECTI Transactions on Computer and Information Technology, vol. 18, no. 1, 10 Feb. 2024, pp. 89–100,https://doi.org/10.37936/ecti-cit.2024181.254501.

[66] Almalaq, Abdulaziz, and George Edwards. "A Review of Deep Learning Methods Applied on Load Forecasting." *IEEE Xplore*, 1 Dec. 2017, https://ieeexplore.ieee.org/document/8260682.

[67] Z. Zhang, "Improved Adam Optimizer for Deep Neural Networks," *2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS)*, Jun. 2018, doi: https://doi.org/10.1109/iwqos.2018.8624183.

[68] Pan, Yan, and Yuanzhi Li. "Toward Understanding Why Adam Converges Faster than SGD for Transformers." *ArXiv.org*, 31 May 2023, https://arxiv.org/abs/2306.00204.

[69] Wilson, D.R., and T.R. Martinez. "The Need for Small Learning Rates on Large Problems." *IEEE Xplore*, 1 July 2001, https://ieeexplore.ieee.org/abstract/document/939002.

[70] Dey, Nilanjan, et al. "Customized VGG19 Architecture for Pneumonia Detection in Chest X-Rays." *Pattern Recognition Letters*, vol. 143, 1 Mar. 2021, pp. 67–74, www.sciencedirect.com/science/article/abs/pii/S0167865520304414, https://doi.org/10.1016/j.patrec.2020.12.010. Accessed 26 May 2021.

[71] M. Sandler, A. W. Howard, M. Zhu, Andrey Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *arXiv (Cornell University)*, Jan. 2018, doi: https://doi.org/10.48550/arxiv.1801.04381.

[72] Z. Rahman, Md. S. Hossain, Md. R. Islam, Md. M. Hasan, and R. A. Hridhee, "An approach for multiclass skin lesion classification based on ensemble learning," *Informatics in Medicine Unlocked*, vol. 25, p. 100659, 2021, doi: https://doi.org/10.1016/j.imu.2021.100659.

[73] Romero-Lopez, Adria, et al. "Skin Lesion Classification from Dermoscopic Images Using Deep Learning Techniques." *Biomedical Engineering*, 2017, https://doi.org/10.2316/p.2017.852-053. Accessed 9 Apr. 2022.

[74] C. Barata, J. S. Marques and M. E. Celebi, "Deep Attention Model for the Hierarchical Diagnosis of Skin Lesions," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, CA, USA, 2019, pp. 2757-2765 doi: 10.1109/CVPRW.2019.00334.

[75] Gessert, Nils, et al. "Skin Lesion Diagnosis Using Ensembles, Unscaled Multi-Crop Evaluation and Loss Weighting." *ArXiv:1808.01694 [Cs]*, 5 Aug. 2018,

arxiv.org/abs/1808.01694.

[76] Mahbod, Amirreza, et al. "Transfer Learning Using a Multi-Scale and Multi-Network Ensemble for Skin Lesion Classification." *Computer Methods and Programs in Biomedicine*, vol. 193, Sept. 2020, p. 105475, https://doi.org/10.1016/j.cmpb.2020.105475. Accessed 25 Nov. 2020.

[77] Ahmed, Sara Atito Ali, et al. "Skin Lesion Classification with Deep CNN Ensembles." *IEEE Xplore*, 1 Oct. 2020, ieeexplore.ieee.org/document/9302125.

[78] Steppan, Josef, and Sten Hanke. "Analysis of Skin Lesion Images with Deep Learning." *ArXiv.org*, 11 Jan. 2021, arxiv.org/abs/2101.03814.

[79] Menghani, Gaurav. "Efficient Deep Learning: A Survey on Making Deep Learning Models Smaller, Faster, and Better." *ArXiv:2106.08962 [Cs]*, 21 June 2021, arxiv.org/abs/2106.08962.

[80] Raza, Khalid, and Nripendra Kumar Singh. "A Tour of Unsupervised Deep Learning for Medical Image Analysis." *ArXiv (Cornell University)*, 19 Dec. 2018, https://doi.org/10.48550/arxiv.1812.07715. Accessed 13 Dec. 2023.

[81] Knolle, Moritz, et al."Efficient, High-Performance Semantic Segmentation Using Multi-Scale Feature Extraction." *PLOS ONE*, vol. 16, no. 8, 19 Aug. 2021, p. e0255397, https://doi.org/10.1371/journal.pone.0255397. Accessed 26 Sep. 2021.

[82] Kazim, I.; Gande, T.; Reyher, E. .; Gyatsho Bhutia, K. .; Dhingra, K.; Verma, S. Advancements in Sequencing technologies:: From Genomic Revolution to Single-Cell Insights in Precision Medicine. J. Knowl. Learn. Sci. Technol. 2024, 3 (4), 108-124. https://doi.org/10.60087/jklst.v3.n4.p108.

[83] Ruthwik Guntupalli, Saloni Verma and Karan Dhingra 2024. Impact of Healthcare Digitization: Systems Approach for Integrating Biosensor Devices and Electronic Health with Artificial Intelligence. American Scientific Research Journal for Engineering, Technology, and Sciences. 98, 1 (Aug. 2024), 246–257, https://asrjetsjournal.org/index.php/American_Scientific_Journal/article/view/10786/2789.

[84] Mehta, A.; Alaiashy, O.; Kumar, P.; Tamilinian, V.; Besong, S.; Balpande, S.; Verma, S.; Dhingra, K. Advancing Model-Based Systems Engineering in Biomedical and Aerospace research:: A Comprehensive Review and Future Directions. J. Knowl. Learn. Sci. Technol. 2024, 3 (4), 133-147. https://doi.org/10.60087/jklst.v3.n4.p133.