Preprints.org

Article

Explainable AI for Skin Disease Classification Using Grad-CAM and Transfer Learning to Identify Contours

S M Saiful Islam Badhon , Sharun Akter Khushbu [*] , Noyon Chandra Saha , Asif Hossain Anik , Md. Aiyub Ali , K.S.M. Tozammel Hossain

Article

Explainable AI for Skin Disease Classification Using Grad-CAM and Transfer Learning to Identify Contours

S M Saiful Islam Badhon [1,†] , **Sharun Akter Khushbu** [2,*†], **Noyon Chandra Saha** [2,†], **Md. Aiyub Ali** [2,†], **Asif Hossain Anik** [3,†] **and K.S.M. Tozammel Hossain** [1]

1   University of North Texas; smsaifulislambadhon@my.unt.edu
2   Daffodil International University; sharun.cse@diu.edu.bd
3   Jagannath University; asifhossain2994@gmail.com
*   Correspondence:sharun.cse@diu.edu.bd; Tel.: +880-01730599307
†   These authors contributed equally to this work.

Abstract: This research evaluates the feasibility of addressing computer vision problems with limited resources, particularly in the context of medical data where patient privacy concerns restrict data availability. (1) Background: The study focuses on diagnosing skin diseases using five distinct transfer learning models based on convolutional neural networks. Two versions of the dataset were created, one imbalanced (4092 samples) and the other balanced (5182 samples), using simple data augmentation techniques. Preprocessing techniques were employed to enhance the quality and utility of the data, including image resizing, noise removal, and blur techniques. The performance of each model was assessed using fresh data after preprocessing. According to the research findings, (2) Methods: the VGG-19 model achieved an accuracy of 95.00% on the imbalanced dataset. After applying augmentation on the balanced data, the best-performing model was VGG-16-Aug with an accuracy of 97.07%. (3) Results: These results suggest that low-resource approaches, coupled with preprocessing techniques, can effectively identify skin diseases, particularly when utilizing the VGG-16-Aug model with a balanced dataset. (4) Conclusions: The study addresses rare skin disorders that have received limited attention in past research, including acne, vitiligo, hyperpigmentation, nail psoriasis, and SJS-TEN. The findings highlight the potential of simple data augmentation techniques moreover, explainable AI: Grad-CAM interpreted the model outcome by showing image contours visually as well as identifying uncommon skin conditions and overcoming the data scarcity challenge. The implications of these research findings are significant for the development of machine learning-based diagnostic systems in the medical field. Further investigation is necessary to explore the generalizability of these findings to other medical datasets.

Keywords: skin disease; transfer learning; VGG-16; CNN; explainable AI; Grad-CAM

---

1. Introduction

The multiple body limbs have the combined structure of a human. There are two-part inner sculptures and the outer cover is skin which can protect our pathogens, immunity, excessive water loss, and another concerning part of our body [1]. Skin disease is a common infectious in the universe [2]. The World Health Organization (WHO) also claimed that by 2020 [3] 49.5% of Asian women and men will have lost their lives to skin diseases. Statistics of death also added that 384 people died in Bangladesh [4]. However, this number globally growing to 9.6 million deaths has occurred [5]. When people are getting attacked their skin texture becomes in downslope. The skin disease reasons are viruses, bacteria, allergy, or fungal infections [6]. It happens to people as well as genetically also. Tropically, most of the disease happens in the thinner layer of the skin. This type of outer layer disease contaminates the skin which name is the epidermis besides, visible to others and can be responsible for mental depression and injuries. There are several types of skin sore: Acne, Vitiligo, Hyperpigmentation (HP), Nail-psoriasis (NP), Stevens-Johnson syndrome, and toxic epidermal necrolysis (SJS-TEN) as shown in figure 2. The sore is not similar regarding skin image sign and hardness, Tropically, few diseases are recovered by changes of time, and some of the diseases are fixed and might be painful or not. According to these diseases are awful including SJS-TEN most degrading type. Skin problems

obviously can be cured at an early-stage prediction. Tropically, extensive lack of linkage between medical specialists and affected skin disease people in fact, people are unaware of their health skin types, symptoms, and most concerns related to early-stage prediction. In the Biomedical sector, there are lots of diseases that are invisible signs but it affects the body's inner part. In the sense of this purpose, it has to be a very quick prediction as long as report solution. In practice, there are lots of obstacles to diagnosing and cost-effectiveness to knowing disease type and skin forecasting results. Cross the ML approach Deep learning Convolution Neural Network is the highest possible architecture to detect any images regarding faster and more accurate solutions. Google has a tremendous quantity of studies available online, all are based on skin disease detection or classification for the last few decades. Approach some automatic computer-aided technology from biomedical skin-related images is significant research in the Deep learning area [7]. Moreover, a huge quantity of quality research has been proposed in different ways of experimenting on recognition or classification, however, continuously researchers try to find out the lack to be filled. Recent studies have been available on multilevel disease [8][9] but single images have been done before [10][11]. These studies are inadequate for identifying multiscale classes [12]. The multiscale class assessment is a significant challenge depending on the skin disease similarities such as the same pattern as different diseases. Skin problems obviously can be cured at an early-stage prediction. Tropically, extensive lack of linkage between medical specialists and affected skin disease people- ple in fact, people are unaware of their health skin types, symptoms, and most concerns related to early stage prediction. In the Biomedical sector, there are lots of diseases that exist that are invisible signs but they affect the body's inner parts. In the sense of this purpose, it has to be a very quick prediction as long as report solution. Practically, there are lots of obstacles to diagnosing and cost-effectiveness knowing disease type and skin forecasting results. Cross the ML approach Deep learning Convolution Neural Network is the highest possible architecture to detect any images regarding faster and more accurate solutions. Google has a tremendous quantity of studies available online, all are based on skin disease detection or classification for the last few decades. Approach some automatic computer-aided technology from biomedical skin-related images is significant research in the Deep learning area [7]. Moreover, a huge quantity of quality research has been proposed in different ways of experimenting on recognition or classification, however, continuously researchers try to find out the lack to be filled. Recent studies have been available on multilevel disease [8] [9] but single images have been done before [10] [11]. These studies are inadequate for identifying multiscale classes [12]. The multiscale class assessment is a significant challenge depending on the skin disease similarities such as the same pattern as different diseases. The key contribution of this study is outlined as follows:

- Evaluation of Deep Learning Models: The study comprehensively evaluated five deep learning models, including GoogleNet, Inception, VGG-16, VGG-19, and Xception, for the identification of skin diseases.
- Creation of a Self-Collected Dataset: In this study, a dataset consisting of approximately 5,185 images of rare skin diseases was created.
- Importance of Data Balancing and Augmentation: The study highlighted the significance of data balancing techniques, specifically through augmentation, in improving the accuracy of skin disease identification models. By applying augmentation on the imbalanced dataset, the study successfully enhanced the performance of the models, with VGG-16-Aug achieving the highest accuracy of 97.07% on the balanced data. This emphasizes the importance of addressing imbalanced data distributions for accurate disease prediction.
- We have used lime to explain how the model performs in the hidden layer and also identify the disease areas by using the deep color contrast.

The remaining of the research study is structured as follows: Section 2 provides a detailed background and statement of limitations. In section 3, the methodology is elaborated, which includes pre-processing, segmenting the portion of the skin, augmenting the data, and the crucial aspect of disease recognition. Section 4 presents the practical implementation outcomes, with a discussion

on quantitative evaluation and durability. In section 5 conclusion of the study is discussed, and potential areas for future research are identified. Finally, in section 6, the research study is summarized, including key findings and their implications.

2. Literature Review

In recent research studies, most available datasets have been used to build skin disease detection systems. We followed a different process by using low resources to collect medical images from various sources, in order to improve model performance in five classes. For three classes, we collected less than 1000 images out of the five classes. Due to the low resources, we used several deep techniques to build our own model, which was able to identify images with approximately 98% accuracy. Additionally, our research pursued a state-of-the-art approach by conducting comparative studies. A CNN architecture with the SoftMax classifier was used to detect dermoscopic images, and when checked with ML classifiers, it produced the highest accuracy and generated a diagnostic report as an output [13]. In [14], a prominent study on 56,134 data from 17 clinical instances was conducted. Set A contained 16,539 images, while [15] Set B contained 4,145 images of 26 different skin combinations, resulting in 94% accuracy. In [16], a classification approach assumed 85% accuracy and was implemented with MobileNet, LSTM, CNN, and VGG using the HAM1000 dataset. CNN [17] is the best solution for image recognition, scoring 90% to 99% accuracy with 60 different skin problems. Another study on skin diseases showed the study of VGG16, LeNet-5, and AlexNet on 6,144 learning images with 5 distinct problems, and low resource images were used [21]. MobileNet [19] provided a hybrid loss function with the modified architecture, resulting in 94.76% accuracy. In CNN [20], 2475 dermoscopic images [21] were identified with 88% accuracy, while other classifiers such as GBT, DT, and RF were also used. A new modified loss function has been proposed using DenseNet201 [22], achieving 95.24% accuracy. In [23], a review was conducted on 16,577 images with 114 classes on Fitzpatrick skin type labels, resulting in a profound solution in DNN and better accuracy. In [24], research was conducted on lesion skin medical images, and advances in DCNN were measured by the sensitivity of 90.3% and specificity of 89.9%. Using 1834 images, a DL model [25] was trained to achieve a 95% score and a confidence interval error margin of 86.53%, ROCAUC of 0.9510, and Kappa of 0.7143, with both high sensitivity and specificity. A low resource faster solution was proposed in [26], which trained up medical three types of skin problems with CNN and Multiclass SVM achieving 100% accuracy. Another case study approached DL with 22 different types of skin with higher accuracy [27]. The global dataset ISIC2017 of Melanoma skin diseases was used in a study that used Deeplabv3plus, Inception-ResNet-v2-uNet, MobileNetV2-unet, ResNet50-uNet, and VGG19-uNet. Instead of Deeplabv3plus, the model that showed the highest recall of 91% was chosen, and both preprocessing methods were applied in 5 models. In [28], the same dataset was used from HAM10000 with 100154 images that were trained with ResNet-50, DenseNet-121, and a seven-layer CNN architecture. These models achieved a perception of 99% accuracy on extracting features. In [29], it was claimed that after resizing, augmentation, and normalizing the same data from ISIC2018, models tuned up on CNN, ResNet50, InceptionV3, and ResNet+InceptionV3 combined up models achieved more than 85.5%. In [28], researchers examined 58,457 skin images, including 10,857 unlabeled samples, for multilevel classification. They achieved a perfect AUC score of 97% and a high F1-macro score, which helped to solve the problem of imbalanced images. In [30], a comparison was made between ML classifiers and CNN. The study found that the Deep Convolutional Network suggested the best-tuned architecture for separate image cell prediction. Another study, [31], found that classification for feature extraction learning was capable of categorizing all same-level diseases. The purpose of this study was to experiment on the MIAS dataset using AlexNet in addition to NB, KNN, and SVM. One of the prominent studies, [32] examined pre-stage signs of skin disease using 5 different DL models and achieved a high accuracy of 99% on the HAM10000 dataset. In [33], researchers experimented with EfficientNetV2 to reduce the limitation of acne, actinic keratosis (AK), melanoma, and psoriasis image classification, achieving a result of 87%. In [34], the ISIC dataset was approached with a total of 8917 medical images trained on a

CNN architecture. EfficientNetB5 achieved an accuracy of 86% in identifying pigmented lesions and improved the AUROC curve to 97%. In [35], SCM (Spectral Centroid Magnitude) was suggested and combined with KNN, SVM, ECNN, and CNN to achieve a score of up to 83% on 3100 images from PH2 and ISIC. MobileNetV2 and LSTM [36] combined architecture achieved a score of 86% with high performance and low error occurrence. In contrast, both datasets ISIC and HAM10000 with 5 distinct diseases performed SVM, KNN, and DT, contributing to image preprocessing, segmentation, feature extraction, and classification that outperformed the study.

3. Materials and Methods

This section outlines the proposed methodology for identifying low-resource skin diseases using both augmented and non-augmented approaches. Each evaluated study is described sequentially, starting with image processing, followed by image augmentation, feature extraction, and finally implementation using deep learning techniques. Figure 1 provides an overview of our working procedure.
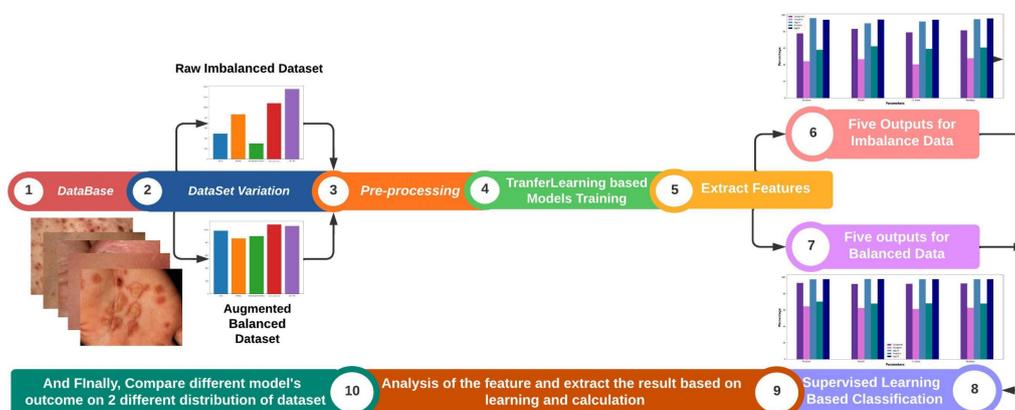
Figure 1. A Proposed Diagram on Skin Disease Classification System shows Dataset with class distribution, image data analysis, feature engineering and final result chart on transfer learning and supervised learning.

3.1. Dataset Description

The biggest contribution and most time-consuming task are preparing a self-collected dataset. However, there are several benefits in biomedical research, such as using newly identified skin images that other researchers have not used in their methodology. On the contrary, the most challenging aspect is that self-collected images are not preprocessed beforehand, making model training quite difficult. In regards to skin disease images, there are various internationally available collections of dermatological images, such as ISIC2019 [37] and HAM10000 [38], which are the two largest datasets for melanoma skin disease. These datasets cover several skin conditions, including actinic keratosis, basal cell carcinoma, benign keratosis, dermatofibroma, melanoma, melanocytic nevus, squamous cell carcinoma, and vascular lesions. However, we have collected images of a few skin diseases that rarely occur in the human body, such as acne, vitiligo, hyperpigmentation, nail psoriasis, and SJS-TEN showed in Figure 2 and total dataset description depicts in Table 1.

Table 1. Dataset Description.

Name	Description
Total number of images	5184
Dimension	224*224
Image format	JPG
Acne	984
Hyperpigmentation	900
Nail psoriasis	1080
SJS-TEN	1356
Vitiligo	864

Figure 2. Sample images of five skin images such as Acne, Hyperpigmentation, Nail psoriasis, Vitiligo, SJS-TEN.

Our dataset[49] consists of 5 categories of skin images, each with detailed depictions of skin sores and injuries in various ways. We collected a total of 5184 dermatoscopic images from different hospitals and online resources of patients from different countries. This indicates that transfer learning can be used to detect all types of skin diseases. Specifically, we worked on the categories of acne (984), vitiligo (864), hyperpigmentation (900), nail-psoriasis (1080), and SJS-TEN (1356).

3.2. Dataset Preparation

There are some challenges in self-collected data, such as a lack of resources and an imbalance in the dataset occurrences. We have five categories: acne, vitiligo, hyper-pigmentation, nail-psoriasis, and SJS- TEN. We collected online sources with imbalanced data, including 492 images of acne, 864 images of vitiligo, 300 images of hyper-pigmentation, 1080 images of nail-psoriasis, and 1356 images of SJS-TEN. We have taken imbalanced data in each class, as the data distribution of acne and hyper-pigmentation image collections is very poor, which is responsible for biased accuracy and ignores the minority of image classes. According to the minority of classes, allowing a high error rate, the prediction will fall.

3.3. Random Oversampling

When one category has a minority of image ratio, oversampling is a technique that can be used to enhance the image quantity by duplicating the image with different angles, rotations, and flipping. In

our dataset, we have two classes with image ratios lower than the others. We applied oversampling only to the acne and hyper-pigmentation classes. As a result, we achieved a balanced dataset, which is shown in Figure 3.

(a)   **(b)**
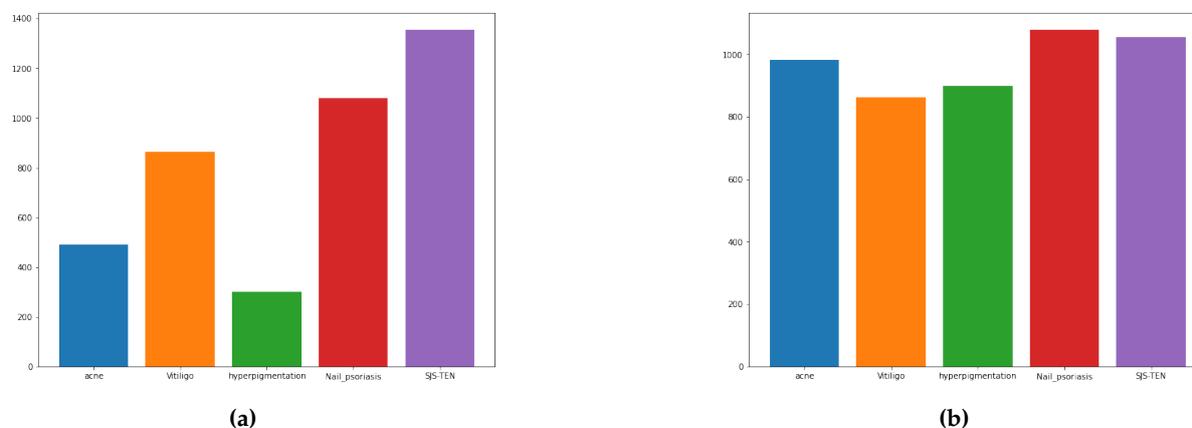
Figure 3. Distribution of Data before (a) and after (b) Random Sampling with Augmentation.

3.4. Image Processing

Before image processing, the sizes, shapes, and colors of the images are not in a usable format. Therefore, raw images need to be preprocessed. In raw skin images, there are countless cracks and distortions. Acquiring high-resolution or good-quality images can significantly improve the detection of skin diseases. Quality images can also optimize complexity, reduce loss curves, and achieve high accuracy. In this subsection, we will discuss the preprocessing steps, including image resizing.

3.5. Image Resizing

Usually, raw collections of images are in different formats, which can lead to imbalanced image features. Technically, the total dataset should be unified into one structure by resizing the image shape. Different sizes of images can be resolved using increasing or decreasing resize matrix operations. There are two specific solutions for effective performance and reduced complexity metrics. In this study, all the input images in different shapes were later converted to a size of 224x224.

3.6. Noise Removal

Noise is an immature fact that occurs as a disturbance when any image is processed, flipped, or generates an output. This causes dependence on removing noise. Instead, these pictures lose their proper size and degrade in quality due to imbalanced color brightness. To apply deep techniques, a few filtering processes appear in digital images to enhance image shape while preserving edges and protecting central edges, and processing 2D images. We can use more than three filters to improve image edges, such as Gaussian blur, averaging blur, median blur, and bilateral blur filtering, which help to blur and remove relevant image noise [39].

3.7. Augmentation

Image augmentation is the process of augmenting the minimum resources of an image into non-duplicate regions. Typically, image augmentation involves the reflection of various types of images on texture, grayscale, low-high brightness, color contrast, and other image features. Augmentation creates bounding boxes to improve accuracy with compilation time, which is known as synthetic data. Throughout the Image augmentation is the process of augmenting the minimum resources of an image into non-duplicate regions. Typically, image augmentation involves the reflection of various types of images on texture, grayscale, low-high brightness, color contrast, and other image features. Augmentation creates bounding boxes to improve accuracy with compilation time, which is known as synthetic data.

3.8. Blur Techniques

Gaussian Blur is the simplest way to blur any images. Using this technique, images can be modified to shape out the actual object as per the user's requirements. Quality and quantity can also be improved through this process. This involves using a mathematical equation on a slightly changed matrix formation to create an image. We also used Averaging Blur, Median Blur, and Bilateral Blur for better augmentation of image quality and quantity. Figure 4 shows samples of blur representing the matrix deformation pattern. In Table 2, we used several blur techniques and included all the equations that we implemented on our low-resource skin images. Regarding our self-collected images, we captured color images. For skin images, we computed different features such as contrast, energy, entropy, correlation, and homogeneity from the image matrix.

3.8.1. Gaussian Blur:

The Gaussian blur technique involves generating an n x m matrix and computing new RGB values for the blurred image. It applies a Gaussian function to each pixel in the image to determine its new value.

$$GB(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{x^2 + y^2}{2\sigma^2}}, \tag{1}$$

3.8.2. Averaging Blur:

The averaging blur technique is used to remove high-frequency content, such as edges, from the image, resulting in a smoother appearance. It works by taking the average of pixel values within a specified neighborhood around each pixel. x[n] = represents the pixel values and k ranges from 0 to n.

$$AB(n) = \frac{1}{M} \sum_{M-1}^{K=0} x[n-k], \tag{2}$$

3.8.3. Median Blur:

The median blur technique replaces each pixel's intensity with the median intensity value from neighboring pixels. This helps in reducing noise and preserving edges in the image where, I (x, y, t) = represents the intensity value of the pixel at coordinates (x, y) and frame t.

$$MB(x,y,t) = median\, I(x,y,t-i), \tag{3}$$

3.8.4. Bilateral Blur:

Bilateral blur replaces the intensity of each pixel with a weighted average of intensity values from nearby pixels. It considers both spatial distance and intensity differences when calculating the weights. This technique preserves edges while smoothing the images where, BF[I]p = represents the blurred intensity value at pixel p, S= Denotes the spatial neighborhood, Ip and Iq = The intensity values of pixels p and q, respectively, and Gσs = The spatial Gaussian function.

$$BF[I]_P = \frac{1}{W_p} \sum_{Q\epsilon S} G\sigma s(p-q) G\sigma r(I_p - I_q) I_q, \tag{4}$$

Figure 4. Image samples of the transferred image into blur image, therefore, several image machines can understand.

Table 2. Random sample images from different skin images where the high score of each measure represents transfer learning on images can influence hidden layer adaptation.

Image	Image_1	Image_2	Image_3	Image_4	Image_5	Image_6	Image_7	Image_8	Image_9	Image_10
Energy	0.0485	0.0290	0.0286	0.0286	0.0325	0.0284	0.0348	0.0636	0.0392	0.0478
Correlation	0.8838	0.7944	0.9763	0.9711	0.9838	0.9029	0.9471	0.9964	0.9752	0.9938
Contrast	1112.3901	546.9877	45.1985	64.6160	22.8364	559.1883	244.3821	1.8534	27.5155	4.4025
Homogeneity	0.3357	0.2438	0.2451	0.3491	0.3242	0.2446	0.4573	0.5935	0.3337	0.4511
Entropy	10.8958	11.5756	11.1457	10.9050	10.7034	11.2315	10.4978	8.3734	10.3353	9.2798
Mean	0.3171	0.3760	0.5281	0.2792	0.4288	0.2504	0.3817	0.3769	0.5094	0.5319
Variance	0.0735	0.0204	0.0146	0.0172	0.0109	0.0443	0.0358	0.0039	0.0086	0.0055
SD	0.2711	0.1429	0.1209	0.1313	0.1046	0.2104	0.1892	0.0628	0.0927	0.0738
RMS	0.4171	0.4023	0.5418	0.3085	0.4413	0.3270	0.4260	0.3821	0.5177	0.5370

3.9. Statistical Features

In terms of image density, the statistical features use the RGB pattern of the image to measure it. These statistical features find out the extracted colors, separate them, and categorize them using Mean, Feature Variance, RMSE, and Standard Deviation.

3.9.1. Energy:

Energy is a measure of the sum of squared elements in the Gray-Level Co-occurrence Matrix (GLCM). It indicates the overall "energy" or magnitude of the pixel relationships in the GLCM. The formula for energy is the sum of squared elements in the GLCM, normalized to a value between 0 and 1. Correlation: Correlation measures the overall intensity of the relationship between a pixel and its neighboring pixels in the image. It quantifies the linear dependency between pixel values in the GLCM. The formula for correlation is the sum of the product of the normalized GLCM elements and the product of their respective deviations from the mean.

3.9.2. Contrast:

Contrast provides a measure of how closely a pixel is connected to its neighbors in terms of intensity differences. It quantifies the local variations or differences in pixel values within the GLCM. The formula for contrast is the sum of the product of the squared difference between gray-level pairs and their respective frequencies in the GLCM.

3.9.3. Homogeneity:

Homogeneity describes the closeness or uniformity of the distribution of elements in the GLCM. It measures how similar or homogeneous the gray-level pairs are in the GLCM. The formula for homogeneity is 1 divided by the sum of the squared differences between gray-level pairs in the GLCM.

3.9.4. Entropy:

Entropy represents the degree of randomness or uncertainty in the distribution of gray-level pairs within the GLCM. It measures the level of information or disorder in the GLCM. The formula for entropy is the sum of the product of the GLCM elements and the logarithm (base 2) of the GLCM element.

Figure 5 displays a collection of image samples that have been converted into blur image matrix representations through a process of image transfer. The features include Energy, Correlation, Contrast, Homogeneity, and Entropy, and are accompanied by their respective formulas and descriptions. These GLCM features are commonly used in texture analysis of images to quantify the spatial relationship between pixel intensities. They provide valuable information for various image processing applications such as pattern recognition, image segmentation, and object detection.
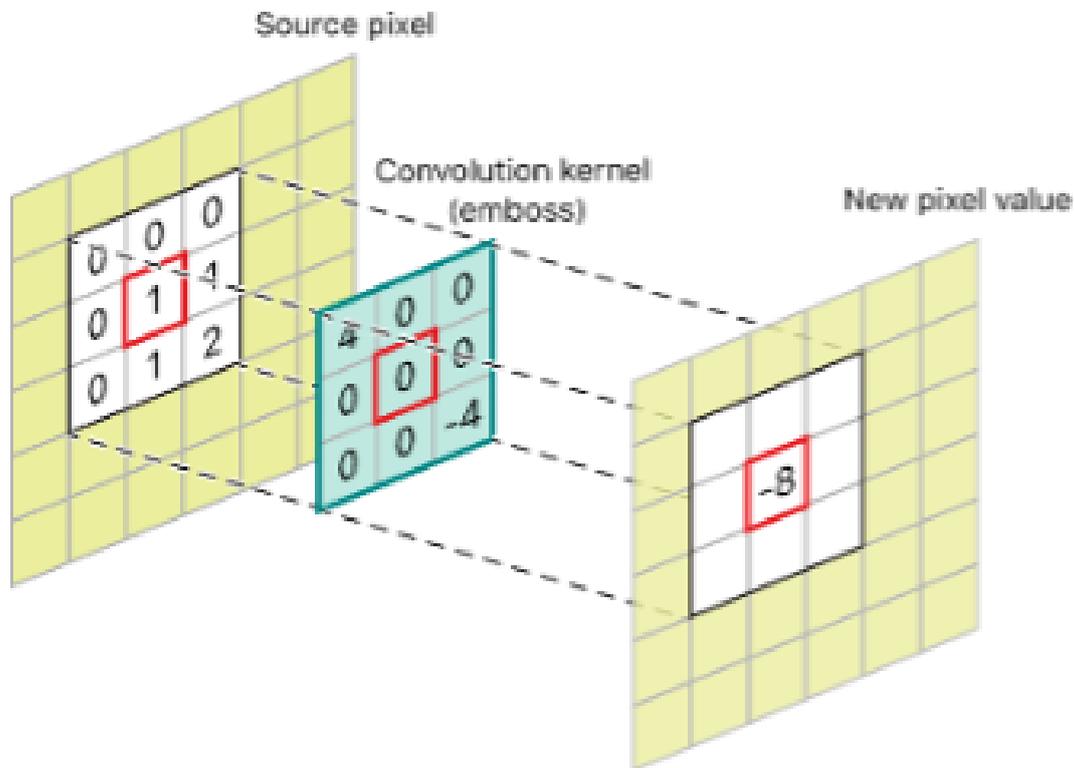

Figure 5. Source pixel to transferred pixel considering ground pixel value.

3.9.5. Mean:

Mean represents the average color value of the image. It is calculated by summing up all the pixel values in the image and dividing it by the total number of pixels where N= total number of pixels in the image and P is the pixel value.

$$Mean = \sum_{i,j=0}^{N-1} iP_{ij}, \tag{5}$$

3.9.6. Variance:

Variance measures the dispersion or spread of the image values around the mean. It quantifies how much the pixel values deviate from the average where μ is the mean value.

$$Variance = \sum_{i,j=0}^{N-1} P_{ij}j(i-\mu)^2, \tag{6}$$

3.9.7. Standard Deviation:

Standard deviation is a statistical measure that quantifies the amount of variability or dispersion within a set of data points. It provides insight into the spread of values around the mean.

$$SD = \sqrt{\sum_{i,j=0}^{N-1} P_{ij}j(i-\mu)^2}, \tag{7}$$

3.9.8. Root Mean Square (Rms):

Root Mean Square is the square root of the average of all squared intensity values in the image. It provides a measure of the overall intensity or energy of the image where, (i - e): Difference between i-th row index and value e is pixel intensity.

$$RMS = \sqrt{\sum_{i,j=0}^{N-1} P_{ij}j(i-e)^2}, \tag{8}$$

3.10. Transfer Learning Based Models

The final step of our work involves classification using transfer learning, which refers to the process of categorizing a group of skin disease data into different classes. Our deep learning approach involves assuming the patterns of skin disease types based on the extracted features of images. For the application development process and self-collected dataset types, we employed multiple models for classification. To the best of our knowledge, we implemented five transformer models: GoogleNet, Inception, VGG-16, VGG-19, and Xception for classifying diseases. GoogleNet [44] is a field of deep learning concerned with various configurations of convolutional neural networks, recurrent neural networks, and deep neural networks used to solve image classification problems, object detection, and feature extraction from skin disease images. It consists of 27 layers with pooling layers. Inception [45] is the most useful architecture for image classification, which uses hidden tricks for gaining output accuracy. This architecture works on a huge variation of images with 3 distinct filter sizes (1x1, 3x3, 5x5) with respect to max pooling layers. Corresponding to these three filter sizes, skin disease image filters are connected in 3 steps, with 1 layer connected to the other layers. VGG-16 [46] uses smaller layers for disease images. Whenever we apply VGG-16 (Visual Geometry Group), it discards the left, right, up, and down images information and focuses on spatial features managed by a 3x3 filter size. VGG-19 (Visual Geometry Group) [47] is trained on 19 layers corresponding to 5 max-pooling and 1 softmax, where it optimizes the output better than Alexnet for disease-related images. VGG-19 works with 5 different skin disease classifiers, producing the best output. Xception [48] is an architecture with 36 layers with spatial features, where disease images are wrapped with 14 modules of linear residual connection.

3.10.1. Custom Cnn Architecture:

Thus, convolution layers used in this work for feature extraction are more than one in the developed private CNN and each of them is followed by the max pooling layer and ReLU activation. In order to extract the aforementioned characteristics, fully connected layers process the final classification output. The structure of the obtaining network as the hierarchy of convolution and pooling with dense layers allows the network to learn and generalize the input properly. Consequently, the network is capable of enforcing the detection of features which are said to be of high level and needful for identifying paddy leaf diseases. The feature extraction in the approach we suggested also heavily depends on this architecture, which was developed carefully to detect patterns from the images of paddy leaves. These patterns are vital in disease identification and classification as explained earlier on in this document. The key components of the custom CNN architecture are comprehensively detailed

as follows:The key components of the custom CNN architecture are comprehensively detailed as follows:

Input Layer : The network takes a scaled image of a skin images with the specifications of 224x224 x 3; 224 being the width and height of the image pixels, and 3 being the number of RGB color channels. This resizing helps the network to handle data consistently and this is wise made by ensuring that the input size is regulated. The raw pixel data is fed to the network at the input layer which can be termed as the initial layer of the neural network.

Convolutional Layers : To obtain the relevant characteristics for recognition, the input image is passed through a number of convolutional layers. Each of them develops feature maps that extract different features such as edges, textures and form by convolving the input with learnable filters (kernels). Mathematically, the convolution operation is defined as:

$$f_k(i, j) = \sum_{m,n} x(i + m, j + n) \cdot w_k(m, n) + b_k, \tag{9}$$

where $x(i + m, j + n)$ are the input pixels that are covered by the filter, $w_k(m, n)$ are the filter weights and b_k is the bias for the kth filter. In this case, it is moved across the input image, and by calculating the filter at each point the dense of patterns corresponding to the filter is determined.

Activation Function: Activation Function: To include non-linearity into the model the output of each of the con- volutional layers is passed through the ReLU activation function. The definition of ReLU is: The definition of ReLU is:

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

This function helps to keep positive values in the feature map while making all the negative pixel values to be zero. Due to this non-linear transformation, the neural network is able to model complicated characteristics by comprehensively learning the various patterns and correlative factors from the given data set.

Pooling Layers: Between the convolutional layers there are layers known as the max-pooling layers whose purpose is to decrease the dimensionality of the feature maps while preserving critical features. The definition of the max-pooling operation is:The definition of the max-pooling operation is:

$$p_{ij} = \max\{x(i + a, j + b)\}, \tag{11}$$

where p_{ij} is the pooled output at the (i, j), position, and a, b are the dimensions of the pooling window. Choosing the maximum value within each window as it slides over the input feature map is known as max-pooling. This downsampling process reduces spatial dimensions, thereby emphasizing the most important features in each region and reducing computational load and overfitting.

Fully Connected Layers: The convolutional layers' high-level features are interpreted by the fully connected (dense) layers of the CNN's final layers, which generate a fixed-size output. The following represents the fully connected layers:

$$y = \sigma(Wx + b), \tag{12}$$

where x is the input vector from the previous layer, y is the output vector, W is the weight matrix, b is the bias term, and σ is the activation function—typically softmax for classification tasks. These layers combine information from all previous layers, enabling the network to make final predictions based on important traits discovered during pooling and convolutional processes. Accurate classification is

facilitated by a feature vector that is the output of the last fully connected layer. This vector captures the most important features of the input image. The softmax activation function in the output layer converts logits into optional divisions. That gives a confidence score for all classes.

First Convolutional Layer: 32 (3x3) filters has been used, then max pooling and ReLU activation.

In the second convolutional layer there are 64 3x3 filters followed by max pooling (2x2) and rectified linear unit activation..

Third Convolutional Layer: It and uses three by three filters 128, max pooling (two by two), ReLU activation.

3.11. Experimental Setup

Our overall experiment was executed in Python language, and for coding, we used Google Colab to extend RAM facilities. We conducted our experiment in two different ways, before and after augmentation. Before augmentation, we used 4092 data in our 5 transformer learning models, and after augmentation, we used 5184 images for implementation, which improved the results of our experimental approach. To ensure a perfectly balanced dataset, we completed preprocessing steps, including resizing all collected images to a 224X224 matrix. This optimized the compilation time and improved the experiment results of the models. Secondly, we implemented a deep technique on an imbalanced dataset, and finally, we applied an augmentation process using Gaussian blur to remove noise and create blur in the images with a 5X5 kernel evaluated for sigma value.

4. Result Analysis & Discussion

In the section on Results and Analysis, we evaluated our performances. Hence, we experimented with our data when it was imbalanced as well as after augmentation on the balanced dataset. The Table 3 provides an overview of the performance metrics for several models: GoogleNet, Inception, VGG-16, VGG-19, and Xception. These metrics include accuracy, precision, recall, and F1-score. Analyzing the table, it is evident that there are notable variations in performance among the models.

Table 3. Performance Matrix of DL Models for Imbalanced Dataset.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
GoogleNet	82.91	90.69	96.61	91.40
Inception	47.40	67.74	88.59	57.87
VGG-16	93.97	92.81	99.30	97.87
VGG-19	95.00	97.72	98.30	99.14
Xception	60.30	68.51	77.60	69.78

Starting with the best performers, VGG-19 stands out with a remarkable accuracy of 95.00%. This model demonstrates high precision (97.72%), indicating a low false positive rate, and impressive recall (98.30%), highlighting its ability to identify positive instances effectively. The F1-score (99.14%) further solidifies VGG-19 as a top-performing model, reflecting a harmonious balance between precision and recall. VGG-16 also performs exceptionally well, achieving a high accuracy of 93.97% and an impressive F1-score of 97.87%. On the other end of the spectrum, Inception exhibits the lowest accuracy among the models at 47.40%. This indicates a significant disparity between the predicted labels and the actual labels, suggesting a considerable number of misclassifications. While Inception shows a relatively higher precision of 67.74%, it suffers from a lower recall of 88.59%, indicating a higher rate of false negatives. Consequently, the F1-score for Inception is comparatively lower at 57.87%, emphasizing the model's struggle to achieve a balance between precision and recall.

The Table 4 presents the performance metrics for several DL models when evaluated on a balanced dataset. These metrics include accuracy, precision, recall, and F1-score. Analyzing the table, we can observe significant variations in the performance of the different models. Starting with the best performers, both VGG-16 and VGG-19 demonstrate exceptional results across the metrics. VGG-16 achieves an impressive accuracy of 97.07%, indicating its ability to correctly classify a high proportion of instances. The precision value for VGG-16 is also high at 98.52%, highlighting its low false positive rate. Additionally, the model shows excellent recall (98.34%), indicating its capability to identify positive instances effectively. The F1-score (99.16%) further solidifies VGG-16 as a top-performing model, indicating a well-balanced trade-off between precision and recall. VGG-19 performs similarly well, with a slightly lower accuracy of 96.29% but maintaining high precision (97.12%), recall (97.82%), and F1-score (98.37%). On the other end of the spectrum, Inception exhibits the lowest performance among the models across various metrics. It achieves an accuracy of 58.20%, suggesting a substantial number of misclassifications. Although Inception demonstrates a precision of 78.72%, indicating a relatively low false positive rate, its recall of 85.83% suggests a higher rate of false negatives. Consequently, the F1-score for Inception stands at 76.05%, reflecting its struggle to achieve a balanced performance between precision and recall.

Table 4. Performance Matrix of DL Models for Balanced Dataset.

| Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
GoogleNet	91.11	97.56	98.87	96.00
Inception	58.20	78.72	85.83	76.05
VGG-16	97.07	98.52	98.34	99.16
VGG-19	96.29	97.12	97.82	98.37
Xception	68.36	73.58	76.47	76.57

Figure 6 showcases the predicted output of skin diseases based on images. The image output provides a visual representation of the model's accuracy in classifying skin diseases, making it easier to identify and treat skin conditions accurately. This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation as well as the experimental conclusions that can be drawn.

Figure 6. Predicted Output of Skin Diseases: This figure illustrates the predicted classifications of skin diseases from images processed by the model.

Explanation Result:

Figure 7 shows the Grad-CAM (Gradient-weighted Class Activation Mapping) visualization for a skin disease as detected by CNN model. The Grad-CAM visualization helps understand which parts of the image are most influential in the CNN model's classification process. Let's break down what this visualization indicates: original image (Leftmost Panel) of the disease without any overlay. Grad-CAM Heatmap (Middle Panel) represents the regions that the CNN model considers significant for identifying the disease. Colors towards the red end of the spectrum (red, orange, yellow) indicate higher importance, while colors towards the blue end (blue, green) indicate lower importance. The central red region shows the most significant area that the model used for classification. Overlay Image (Rightmost Panel) shows original image with the Grad-CAM heatmap overlaid on it. It visually merges the significance of regions (from the heatmap) with the actual insect.
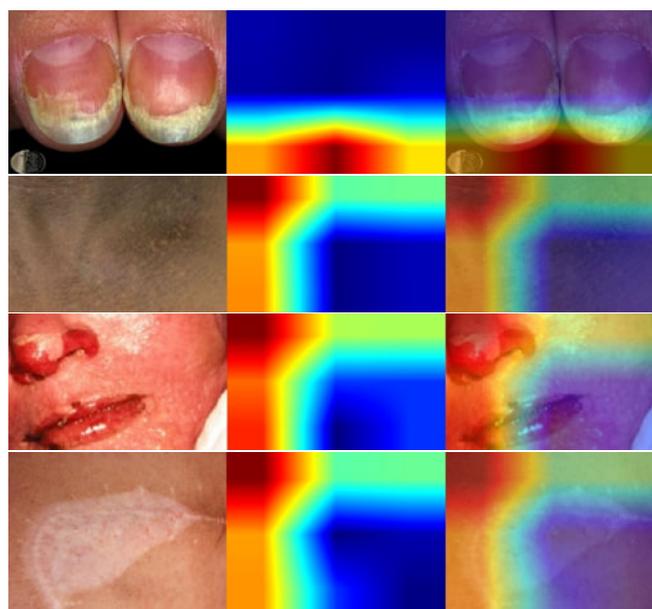
Figure 7. Explanation of Skin Diseases using Grad-CAM Visualization: This figure illustrates the Grad-CAM visualization for a skin disease detected by the model.

5. Conclusion and Future Work

This study aimed to develop an early and predictive system for identifying patterns in skin diseases using advanced deep-learning models. By conducting multiple phases of experimentation and evaluation, we assessed the performance of five prominent models: GoogleNet, Inception, VGG-16, VGG-19, and Xception. Utilizing a self-collected dataset from various online sources, we focused on five skin disease classes, despite the inherent challenge of dataset imbalance. The study underscored the efficacy of data augmentation techniques in addressing class imbalance and enhancing model performance. Notably, VGG-19 achieved an accuracy of 95.00% on the imbalanced dataset, proving its robustness in identifying and classifying skin diseases. Conversely, Inception, with an accuracy of 47.40%, highlighted the limitations of certain models under imbalanced conditions. Upon applying data augmentation, VGG-16 emerged as the best performer with an accuracy of 97.07%, illustrating the critical role of balanced data in achieving reliable and accurate results. The comprehensive statistical evaluations, including Confusion Matrix, Sensitivity, Specificity scores, and validation graphs, further validated the superior performance of our developed models over existing methods. This study's findings emphasize the potential of deep learning models in early skin disease diagnosis, paving the way for their practical application in medical fields. Despite the promising results, this study has several limitations that open avenues for future research. The generalizability of our findings to other medical conditions and datasets remains uncertain. Future studies should explore the application of our methodologies to a broader spectrum of medical domains, encompassing diverse diseases

and disorders. Lastly, further optimization of image preprocessing techniques such as resizing, noise removal, and blur techniques is essential. Future research can explore novel algorithms and approaches to enhance the quality and utility of medical images for diagnosis, thereby improving the overall effectiveness of predictive systems in medical applications.

Author Contributions: Conceptualization, methodology, writing—original draft preparation, formal analysis, Badhon, S. and Khushbu, S.; software, writing—review and editing, Data curation Saha, N.; validation, writing—review and editing, Ali, A., Anik, A., Hossain, T.; funding acquisition Hossain, supervision, Hossain, T.;

Informed Consent Statement: We affirm that this manuscript is unique, unpublished, and not under consideration for publication elsewhere. We confirm that the manuscript has been read and approved by all named authors and that there are no other people who meet the criteria for authorship but are not listed. We further affirm that we have all approved the order of authors listed in the manuscript. We understand that the corresponding author is the sole contact for the editorial process. She is responsible for communicating with the other authors about progress, submissions of revisions, and final approval of proofs. Moreover, this study do not need any consent from patients, as we have collected from online resources.

Data Availability Statement: Data will be made available on request

Conflicts of Interest: Declare conflicts of interest or state "The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.".

Abbreviations

The following abbreviations are used in this manuscript:

DL	Deep Learning
Grad-CAM	Gradient-weighted Class Activation Mapping
VGG-16	Visual Geometry Group 16
SJS-TEN	Stevens-Johnson syndrome and toxic epidermal necrolysis

References

1. G. Martin, S. Gu´erard, M.-M. R. Fortin, D. Rusu, J. Soucy, P. E. Poubelle, and R. Pouliot, Pathologicalcrosstalk in vitro between t lymphocytes and lesional keratinocytes in psoriasis: necessity of direct cell-to-cell contact, *Laboratory investigation*, **2012**, *vol. 92, no. 7*, pp. 1058–1070.
2. R. J. Hay, N. E. Johns, H. C. Williams, I. W. Bolliger, R. P. Dellavalle, D. J. Margolis, R. Marks, L. Naldi, M. A. Weinstock, S. K. Wulf, et al., The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions, *Journal of Investigative Dermatology*, **2014**, *vol. 134, no. 6*, pp. 1527–1534.
3. H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA: a cancer journal for clinicians*, **2021**, *vol. 71, no. 3*, pp. 209–249.
4. M. F. Akter, S. S. Sathi, S. Mitu, and M. O. Ullah, Lifestyle and heritability effects on cancer in bangladesh: an application of cox proportional hazards model, *Asian Journal of Medical and Biological Research*, **2021**, *vol. 7, no. 1*, pp. 82–89.
5. J. D. Khoury, E. Solary, O. Abla, Y. Akkari, R. Alaggio, J. F. Apperley, R. Bejar, E. Berti, L. Busque, J. K. Chan, et al., The 5th edition of the world health organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms, *Leukemia*, **2022**, *vol. 36, no. 7*, pp. 1703–1719.
6. W. Roberts, Air pollution and skin disorders, *International Journal of Women's Dermatology*, **2021**, *vol. 7, no. 1*, pp. 91–97.
7. S. S. Han, M. S. Kim, W. Lim, G. H. Park, I. Park, and S. E. Chang, Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm, *Journal of Investigative Dermatology*, **2018**, *vol. 138, no. 7*, pp. 1529–1538.

8. P. Tschandl, C. Rosendahl, and H. Kittler, The ham10000 dataset, a large collection of multi-source dermato-scopic images of common pigmented skin lesions, *Scientific data,* **2018**, *vol. 5, no. 1*, pp. 1–9.

9. M. H. Jafari, E. Nasr-Esfahani, N. Karimi, S. R. Soroushmehr, S. Samavi, and K. Najarian, Extraction of skin lesions from non-dermoscopic images for surgical excision of melanoma, *International journal of computer assisted radiology and surgery,* **2017**, *vol. 12*, pp. 1021–1030.

10. X. Zhang, S. Wang, J. Liu, and C. Tao, Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge, *BMC medical informatics and decision making,* **2018**, *vol. 18, no. 2*, pp. 69–76.

11. B. Harangi, A. Baran, and A. Hajdu, Classification of skin lesions using an ensemble of deep neural networks, In Proceedings of the 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp. 2575–2578, IEEE, 2018.

12. Z. Ge, S. Demyanov, R. Chakravorty, A. Bowling, and R. Garnavi, Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images, In Proceedings of the International conference on medical image computing and computer-assisted intervention, pp. 250–258, Springer, 2017.

13. S. Inthiyaz, B. R. Altahan, S. H. Ahammad, V. Rajesh, R. R. Kalangi, L. K. Smirani, M. A. Hossain, and A. N. Z. Rashed, Skin disease detection using deep learning, *Advances in Engineering Software,* **2023**, *vol. 175*, p. 103361.

14. Y. Liu, A. Jain, C. Eng, D. H. Way, K. Lee, P. Bui, K. Kanada, G. de Oliveira Marinho, J. Gallegos, S. Gabriele, et al., A deep learning system for differential diagnosis of skin diseases,*Nature medicine,* **2020**, *vol. 26, no. 6*, pp. 900–908.

15. P. N. Srinivasu, J. G. SivaSai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, Classification of skin disease using deep learning neural networks with mobilenet v2 and lstm, *Sensors,* **2021**, *vol. 21, no. 8*, p. 2852.

16. T. Shanthi, R. Sabeenian, and R. Anand, Automatic diagnosis of skin diseases using convolution neural network, *Microprocessors and Microsystems,* **2020**, *vol. 76*, p. 103074.

17. M. Chen, P. Zhou, D. Wu, L. Hu, M. M. Hassan, and A. Alamri, Ai-skin: Skin disease recognition based on self-learning and wide data collection through a closed-loop framework, *Information Fusion,* **2020**, *vol. 54*, pp. 1–9.

21. E. Goceri, Diagnosis of skin diseases in the era of deep learning and mobile technology, *Computers in Biology and Medicine,* **2021**, *vol. 134*, p. 104458.

19. A. Esteva, K. Chou, S. Yeung, N. Naik, A. Madani, A. Mottaghi, Y. Liu, E. Topol, J. Dean, and R. Socher, Deep learning-enabled medical computer vision, *NPJ digital medicine,* **2021**, *vol. 4, no. 1*, pp. 1–9.

20. V. R. Allugunti, A machine learning model for skin disease classification using convolution neural network, *International Journal of Computing, Programming and Database Management,* **2022**, *vol. 3, no. 1*, pp. 141–147.

21. E. Goceri, Deep learning based classification of facial dermatological disorders, *Computers in Biology and Medicine,* **2021**, *vol. 128*, p. 104118.

22. M. Groh, C. Harris, L. Soenksen, F. Lau, R. Han, A. Kim, A. Koochek, and O. Badri, Evaluating deep neural networks trained on clinical images in dermatology with the fitzpatrick 17k dataset, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1820– 1828, 2021.

23. L. R. Soenksen, T. Kassis, S. T. Conover, B. Marti-Fuster, J. S. Birkenfeld, J. Tucker-Schwartz, A. Naseem, R. R. Stavert, C. C. Kim, M. M. Senna, et al., Using deep learning for dermatologistlevel detection of suspicious pigmented skin lesions from wide-field images, *Science Translational Medicine,* **2021**, *vol. 13, no. 581*, p. eabb3652.

24. P. M. Burlina, N. J. Joshi, E. Ng, S. D. Billings, A. W. Rebman, and J. N. Aucott, Automated detection of erythema migrans and other confounding skin lesions via deep learning, *Science Translational Medicine,* **2019**, *vol. 105*, pp. 151–156.

25. M. A. Al-Masni, D.-H. Kim, and T.-S. Kim, Multiple skin lesions diagnostics via integrated deep convolutional networks for segmentation and classification, *Computer methods and programs in biomedicine,* **2020**, *vol. 190*, p. 105351.

26. N. Janbi, R. Mehmood, I. Katib, A. Albeshri, J. M. Corchado, and T. Yigitcanlar, Imtidad: A reference architecture and a case study on developing distributed ai services for skin disease diagnosis over cloud, fog and edge, *Sensors,* **2022**, *vol. 22, no. 5*, p. 1854.

27. M. Abbas, M. Imran, A. Majid, and N. Ahmad, Skin diseases diagnosis system based on machine learning, *Journal of Computing & Biomedical Informatics,* **2022**, *vol. 4, no. 01*, pp. 37–53.

28. F. Weng, Y. Xu, Y. Ma, J. Sun, S. Shan, Q. Li, J. Zhu, and Y. Wang, An interpretable imbalanced semi-supervised deep learning framework for improving differential diagnosis of skin diseases, *arXiv preprint arXiv:2211.10858, 2022.*

29. 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 MDPI* **2022***, vol. 10,* p. 1183.

30. B. Shetty, R. Fernandes, A. P. Rodrigues, R. Chengoden, S. Bhattacharya, and K. Lakshmanna, Skin lesion classification of dermoscopic images using machine learning and convolutional neural network, *Scientific Reports* **2022***, vol. 12, no. 1,* p. 18134.

31. V. D. P. Jasti, A. S. Zamani, K. Arumugam, M. Naved, H. Pallathadka, F. Sammy, A. Raghuvanshi, and K. Kaliyaperumal, Computational technique based on machine learning and image processing for medical image analysis of breast cancer diagnosis, *Security and communication networks* **2022***, vol. 2022,* pp. 1–7.

32. S. N. Almuayqil, S. Abd El-Ghany, and M. Elmogy, Computer-aided diagnosis for early signs of skin diseases using multi types feature fusion based on a hybrid deep learning model, *Electronics* **2022***, vol. 11, no. 23,* p. 4009.

33. R. Karthik, T. S. Vaichole, S. K. Kulkarni, O. Yadav, and F. Khan, Eff2net: An efficient channel attention-based convolutional neural network for skin disease classification, *Biomedical Signal Processing and Control* **2022***, vol. 73,* p. 103406.

34. A. C. Foahom Gouabou, J. Collenne, J. Monnier, R. Iguernaissi, J.-L. Damoiseaux, A. Moudafi, and D. Merad, Computer aided diagnosis of melanoma using deep neural networks and game theory: Application on dermoscopic images of skin lesions, *International Journal of Molecular Sciences* **2022***, vol. 23, no. 22,* p. 13838.

35. K. Sreekala, N. Rajkumar, R. Sugumar, K. Sagar, R. Shobarani, K. P. Krishnamoorthy, A. Saini, H. Palivela, and A. Yeshitla, Skin diseases classification using hybrid ai based localization approach, *Computational Intelligence and Neuroscience,* **2022***, vol. 2022.*

36. P. R. Kshirsagar, H. Manoharan, S. Shitharth, A. M. Alshareef, N. Albishry, and P. K. Balachandran, Deep learning approaches for prognosis of automated skin disease, *Life,* **2022***, vol. 12, no. 3,* p. 426.

37. 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,* **2022***, vol. 75,* p. 102305.

38. F. Alenezi, A. Armghan, and K. Polat, Wavelet transform based deep residual neural network and relu based extreme learning machine for skin lesion classification, *Expert Systems with Applications,* **2023***, vol. 213,* p. 119064.

39. S. Singh et al., Image filtration in python using opencv, *Turkish Journal of Computer and Math- ematics Education (TURCOMAT),* **2021***, vol. 12, no. 6,* pp. 5136–5143.

40. M. Xu, S. Yoon, A. Fuentes, and D. S. Park, A comprehensive survey of image augmentation techniques for deep learning, *Pattern Recognition,* **2023***,* p. 109347.

41. P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, A review of medical image data augmentation techniques for deep learning applications, *Journal of Medical Imaging and Radiation Oncology,* **2021***, vol. 65, no. 5,* pp. 545–563.

42. K. Maharana, S. Mondal, and B. Nemade, A review: Data pre-processing and data augmentation techniques, *Global Transitions Proceedings,* **2022***.*

43. S. Bang, F. Baek, S. Park, W. Kim, and H. Kim, Image augmentation to improve construction resource detection using generative adversarial networks, cut-and-paste, and image transformation techniques, *Automation in Construction,* **2020***, vol. 115,* p. 103198.

44. A. Singla, L. Yuan, and T. Ebrahimi, Food/non-food image classification and food categorization using pre-trained googlenet model, In Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management, pp. 3–11, 2016.

45. Q. Mi, J. Keung, Y. Xiao, S. Mensah, and X. Mei, An inception architecture-based model for improving code readability classification, In Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018, pp. 139–144, 2018.

46. A. Younis, L. Qiang, C. O. Nyatega, M. J. Adamu, and H. B. Kawuwa, Brain tumor analysis using deep learning and vgg-16 ensembling learning approaches, *Applied Sciences,* **2022***, vol. 12, no. 14,* p. 7282.

47. T. Subetha, R. Khilar, and M. S. Christo, Withdrawn: A comparative analysis on plant pathology classification using deep learning architecture–resnet and vgg19, **2021***.*

48. F. Chollet, Xception: Deep learning with depthwise separable convolutions, In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1251–1258, 2017.

49. khushbu, sharun akter, "Skin Disease Classification Dataset", Mendeley Data, V1, 2024, doi: 10.17632/3hck-gznc67.1