Detection of Skin Rashes using Image Processing and Deep Learning Architectures

By

Eftakhairul Islam
20101390
Sad Md. Tafhim
23141086
Dipak Debnath Arka
20301066
Maliha Mussarrat
20101354
Rakibul Hassan Hredoy
20101357

A Thesis submitted to the Department of Computer Science and Engineering
In partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
Brac University
January 2024

Declaration

It is hereby declared that

1. The Thesis submitted is my/our own original work while completing degree at Brac University.

2. The Thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The Thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

EFTAKHAIRUL

Eftakhairul Islam
20101390

Tafhim

Sad Md. Tafhim
23141086

Arka

Dipak Debnath Arka
20301066

Maliha Mussarrat

Maliha Mussarrat
20101354

Hredoy

Rakibul Hassan Hredoy
20101357

Approval

The thesis/project titled "Detection of Skin Rashes using Image Processing and Deep Learning Architectures" submitted by

1. Eftakhairul Islam (20101390)
2. Sad Md. Tafhim (23141086)
3. Dipak Debnath Arka (20301066)
4. Maliha Mussarrat (20101354)
5. Rakibul Hassan Hredoy (20101357)

Of Fall, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 9, 2024.

Examining Committee:

Supervisor:
(Member)

Dr. Muhammad Iqbal Hossain
Associate Professor
Department of Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

Skin diseases have become a growing issue all over the world, skin rashes is one of the most common among them. This paper focuses on establishing a viable deep-learning approach to distinguish between skin rashes. For this purpose, Deep Neural Network Models, such as CNN, MobileNetV2, ResNet50V2, and DenseNet201 algorithms were implemented and analyzed for suitability in this kind of image dataset. In this research, a synthetic dataset was prepared which consisted of 1381 images of three distinct Skin rash diseases, which were preprocessed using image resizing and dataset augmentation process. Among the implemented algorithms, multiple algorithms gave high accuracy scores, that suggested the high quality of the dataset.The models were also interpreted using an Explainable AI technique named LIME. ResNet50V2 provided the highest accuracy among the implemented algorithms followed by MobileNetV2, indicating the viability of using DNN architectures in skin rash detection.

Keywords: Deep Learning; ResNet; MobileNet; CNN; SkinRash; XAI; LIME;

Table of Contents

Chapter 1

Introduction

The outermost and largest organ in the human body is the skin. Skin has a major impact on our well-being as it operates as a shield to protect the body. The skin has layers such as the dermis and epidermis layers. Apart from that subcutaneous tissues along with vessels and nerves such as blood vessels, lymphatic vessels, nerves, and muscles make the skin a complex arena. Skin's ability to prevent the degradation of epidermal lipids is a crucial aspect of its role as a potent protective barrier [2]. Since the skin remains on the outer side it is subjected to several external factors which can be harmful. Several factors contribute to the skin's degradation. Fungal and bacterial infections, viral infections, as well as allergic reactions and many forms of other fungus are the reason that poses a significant risk to an individual's health and put them in danger.

These skin disorders find a favorable environment in tropical countries like Bangladesh, which increases the risk of progression to chronic states or even the formation of skin cancer [3]. Skin rashes are apparent signs of underlying health problems and can represent a wide range of dermatological diseases [2]. These rashes can indicate allergies, infections, or systemic disorders, therefore it's critical to detect and diagnose them accurately and quickly in clinical practice. Conventional diagnosis is primarily based on patient history and visual inspection, both of which are subjective and prone to human error. Notwithstanding its importance, dermatologists and the general public alike frequently struggle with the complexity and diversity of skin problems. Therefore, the demand for more accurate and objective diagnostic techniques is urgent. Better outcomes and a lower chance of developing cancer are both possible outcomes that can be achieved by the timely and accurate identification of skin diseases. Deep learning and machine learning are two methods used in modern diagnostics that are becoming more and more popular because they can recognize patterns with an accuracy that is higher than that of the human eye, which is only 70% [22]. The urgent need for improved diagnostic techniques is highlighted by the delayed detection of many skin disorders' symptoms. Deep learning techniques are among the affordable options that have the potential to save lives, particularly in Bangladesh's rural areas where there are limited medical resources and a high rate of disease transmission [4].

Combining different image-processing techniques offers a promising new approach to the field of dermatology for identifying complex patterns and features present in dermatological pictures [1]. Simultaneously, the integration of deep learning models provides a means to enhance and automate skin rash identification. These cutting-edge methods aim to provide doctors with reliable instruments that provide prompt and accurate diagnostic assistance. By doing this, they enhance conventional methods, resolving diagnostic difficulties and bringing about a notable advancement in dermatological therapy.

In this thesis paper, we meticulously aim to assess the effectiveness of various deep learning methodologies, with a particular focus on the formidable capabilities of deep neural networks. Specifically, our inquiry focuses on the complex analysis and predictive capabilities of models of deep learning within the context of identifying skin conditions such as drug rash, ringworm, and Pityriasis rosea. This investigation digs into the world of skin healthy skin. Convolutional Neural Network (CNN), MobileNetV2, DenseNet201, and ResNet50V2 architectures are utilized, and these frameworks are applied to our selected dataset thoughtfully and methodically, to conduct an in-depth analysis of the effectiveness of various models. To understand their viability and understanding the workings of the architectures, the XAI technique named LIME has been utilized. The ultimate goal we have set for ourselves is to discover the best architectural solution capable of excelling in the complex field of understanding skin rash disease images.

1.1 Motivation

A wide range of skin conditions that are characterized by alterations in the appearance, texture, or color of the outermost layer of skin are grouped under the broader concept of "rashes." Certain symptoms, such as redness, pimples, itching, or irritation, may be associated with these. They can be brought on by a wide range of factors, such as infections, allergies, irritants, or more fundamental illnesses. A wide variety of causes and conditions can give rise to skin rashes, which can also express themselves in a variety of different places. Furthermore, diseases like rosacea or eczema aggravate the redness, irritation, or flakiness in these parts of the body. Rashes can appear on the arms and legs as a result of contact with materials, laundry detergents, or plants like poison ivy that irritate. Warm, humid circumstances in the vaginal and groin area can foster infections caused by fungi, which can result in rashes similar to infections caused by yeast.

Contrary to what its name suggests, ringworm is an infection caused by fungi that impacts the outermost layer of skin rather than a worm. A spherical or shaped like a ring rash that may be scaly, reddened, and bothersome is its defining feature. Ringworm is known medically as "tinea corporis." [16]. The rash that is characteristic of pityriasis rosea appears all of a sudden and is a rather frequent skin illness that is considered to be harmless. Beginning with a single red or pink patch that is slightly elevated, the rash often appears on the chest, abdomen, or back of the affected individual. Although the precise etiology of pityriasis rosea remains unknown, it is hypothesized to be associated with a viral infection, potentially a variant of the herpes virus. [16]. A skin reaction that is brought on by medication is referred to as a drug rash, which is also known as a drug eruption. The rash may extend to other parts of the body, such as the face, limbs, trunk, or mucous membranes.

Recent research has shown that skin problems impact a considerable section of the world's population and represent a significant danger to their health. These discoveries come from studies undertaken in recent years. Through the utilization of algorithms for deep learning and image processing, the goal is to facilitate the implementation of rapid interventions and medicinal treatments to improve diagnostic accuracy and speed. The problem of enhancing the accuracy and efficiency of forecasts presents a significant challenge in the healthcare sector. Through the utilization of deep learning algorithms and image processing methods, overcoming these challenges can be accomplished through the implementation of one of the most fruitful alternatives.

Artificial intelligence, or AI, is extensively used in a wide range of application domains because it can increase decision-making consistency, efficiency, and accuracy. Even while AI algorithms appear to be quite effective at forecasting and are growing in popularity, the majority of these models are seen by users as opaque black boxes that are difficult to grasp. For example, applications for AI systems that make use of XAI (eXplainable Artificial Intelligence) technology are constantly expanding. It makes clear why the predictive ability of deep learning models is equally as important as their precision in generating particular forecasts. Because it guarantees the decision-making model's credibility and openness. These days, deep learning and ensemble models in particular are so complicated that even specialists find it difficult to understand them. Interpretation and preciseness are very different from one another. To address this difficulty, XAI is suggested and used to increase the transparency of AI systems. [28].

When it comes to the diagnosis of skin rashes, image processing is of tremendous assistance since it enables a more precise evaluation of skin disorders and makes it possible to visualize this information more clearly. The implementation of this technology can be accomplished at a cost that is appropriate by making use of standard picture devices such as digital cameras or mobile phones. Using computerized methods, it is possible to conduct analysis that is both quick and accurate, resulting in evaluations that are both consistent and objective.

In today's world, there are a lot of comprehensible patterns, and the vast majority of them are associated with agnostic and local structures. Visualization techniques, information extraction, influence techniques, and example-based explanations are examples of agnostic models [28]. The relevance of the features in the ML model has been determined by utilizing the Shapley value and the SHAP approach.

These days, the most important components in healthcare for decision-making and accuracy prediction are AI and ML approaches. Yet the majority of ML models in use today are transparent. In the medical field, a simple "yes" or "no" response isn't always enough. In certain cases, inquiries such as "where" or "how" have greater bearing [27].

By constructing a local linear regression model around the prediction, A precise explanation of the predictions made by any classifier is provided by LIME. A homogenous distribution of random memory is the sampling process used by LIME. Despite its simplicity, this technique is inefficient since it does not take into account feature correlation [26].

1.2 Research Objective

When it comes to understanding commonalities in data that are visually intriguing, Artificial Intelligence (AI) relies primarily on deep learning models. These models make use of complex neural network designs. Within the realm of artificial intelligence, these models are instances of what is referred to as a subset. CNNs, which stand for convolutional neural networks, are considered to be one of the most important aspects of image processing. It is proficient in the application of layers of convolution to recognize intricate patterns and frameworks inside image files. Complementary architectures such as ResNet and MobileNet can enable customized changes. When it comes to decoding and analyzing data that is visually appealing, these technologies have revolutionized the field of image analysis with their unparalleled precision and efficiency. Possessing the ability to derive structured representation from unprocessed pixels is essential to accomplishing full-picture understanding and analysis.

The purpose of this investigation is to determine the algorithm that is used for deep learning that functions the most effectively in identifying skin rash problems. The MobileNet, ResNet50, and CNN algorithms for skin rash prediction are compared in this study to determine which one has the highest accuracy scores.

To come up with an explanation for a particular occurrence, LIME is a framework that was established not too long ago and may be utilized with any BlackBox categorization model [15]. Within the framework of the XAI architecture, we are utilizing LIME. We can observe the characteristics that are taken into account by our models by utilizing LIME. According to the results of the execution, it is clear that the machine seems to prioritize particular characteristics for each class to determine whether or not such characteristics are truly relevant in that particular class. Firstly, it determines whether or not certain characteristics are present within our classes, after which it expands on other characteristics based on how dependent they are on the aspects that have been detected. The purpose of this procedure is to explain how certain characteristics contribute to the classification, as well as determine whether or not those characteristics coincide with the contribution that was anticipated.

1.3 Thesis's Organizational Structure

The following is the breakdown of the report's formatting: In Chapter 2, the concepts related to the paper's subject matter will be described, and additionally, a review of the studies that came before one. Also, the previously created models will be explained in the background section. In Chapter 3, We shall describe the preparation of the dataset along with the structure of the model. Next, in Chapter 4, the results of the method will be discussed and interpreted. Lastly, In Chapter 5, the conclusion of the paper will be discussed, along with the limitations of the paper.

Chapter 2

Literature Review

2.1 Existing Works Related to Our Research

In [12] the author proposed a very simple, yet very fast and expensive equipment to detect the type of skin diseases. This system works by first obtaining a digital image of the area of the skin that is damaged, after that, resizing the image so that a convolutional neural network can extract characteristics from it that have been trained in advance, and finally classifying the features using a multiclass support vector machine. For their study, they took approximately 100 images, of which twenty were used for validation and the remaining eighty were used for training purposes. The system also contains a preprocessing module, which entails scaling the input image to a set size of 227x227 pixels to ensure consistency in the number of features extracted from each image. This is accomplished by resizing the image using the same number of pixels on each side. The AlexNet model, which has already been pre-trained using the ImageNet dataset, was chosen to serve as the CNN for this investigation. To classify the collected characteristics into one of three categories—melanoma, eczema, or psoriasis—the support vector machine (SVM) is utilized. In conclusion, the system that was developed was effectively capable of diagnosing three distinct skin illnesses with a success rate of 100%. In addition to that, they allowed some room for future enhancements to the system.

The model that has been suggested and is provided in this paper was evaluated using three machine learning algorithms: Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbor (K-NN) classifiers [19].The focus of this model was on a total of four different skin disorders. Several image-processing processes are involved in creating this model. They are visualizing the processes of acquisition, preprocessing (including resizing, color alteration, de-noising, and normalization), classification, and feature extraction. For the research around 377 images are taken. In the phase related to the pre-processing of the data, the images were scaled down to 250 by 250 pixels. To get rid of the noises in the skins, a median filter denoising technique was utilized. The Gabor and Sobel procedures were utilized concurrently to extract texture and edge features for feature extraction. According to the findings, The SVM classifier had the best accuracy, which was 90.7%; the RF classifier came in second with 84.2%, and the K-NN classifier came in third with 67.1%. The authors compared their proposed model with previous research that tested some of the diseases studied in this work. When it comes to the accuracy rates for the four skin diseases that were chosen, SVM performs significantly better than any of the other two classifiers. When evaluated with the RF classifier, acne had the highest percentage of accuracy. In comparison, 5 K-NN yields the results that are the least accurate across the board for all of these disorders. They concluded that the studies that were being compared to the suggested model had a lower degree of precision than the proposed model. Additionally, it is mentioned by the authors that cherry angioma disease has been checked for relatively infrequently in earlier studies. Cherry angioma disease was examined using the proposed model, and the findings demonstrate that the model attained a diagnostic accuracy rate of 100% when attempting to diagnose this condition.

To solve the difficulties that are connected with skin lesion categorization, the author of this research suggested a CNN-based model that makes effective use of kernels and activation Functions[20]. His goal was to achieve an extremely high level of accuracy with this classification. The HAM10000 data set, which is considered the gold standard for skin lesions, was utilized for this study. It includes 10,015 images of skin lesions belonging to various populations and is organized into seven main categories, including melanocytic nevi, basal cell carcinoma, melanoma, vascular lesions, benign keratosis-like lesions, actinic keratoses, and dermatofibroma. The best possible outcomes can be accomplished by using Dynamic-sized kernels in the layers, which led to an extremely low number of trainable parameters. When taking into consideration the fact that they took pictures from HAM10000, not all of the images were the same size. However, to train a model using deep learning effectively, you will require data that is acceptable and balanced. To get around this, the appropriate pre-processing was performed, such as imagining scaling to avoid the model from being trained with a biased set of data and augmentation to increase the robustness of the trained model. Both of these steps were taken to overcome this obstacle. They employed the ReLU activation function to reduce linearity, and they used LeakyReLU to reduce the chance of significant information being lost due to the features being represented in the feature maps. Both of these techniques are related to the ReLU activation function. On the validation dataset, the suggested model achieved a great class-wise accuracy (seven classes), as well as an overall accuracy of 97.85 percent.

Lentigo maligna, superficial spreading, and nodular melanoma are distinguished using the author's innovative technique in this article [21]. They employed CNN, the most popular and effective Deep Learning algorithm, to process photos. After delving further into the network, the layers will extract more precise higher-level features. To find the melanoma, they employed two classifications, correct and wrong. The steps they took to fit into the system that will be put into place. a) Improve the image's quality by removing any hairs. b) After extracting several features from the input picture dataset, a training file is created. c) The CNN classification method is used to classify both the recently produced test input pictures and the newly assembled training file collection. They employed a 2475 picture dataset obtained from the "dermnetNZ" website, which has 783, 936, and 756 photographs of malignant, superficial spreading, and nodular melanoma, respectively. To avoid the data being overfitting, they adopted a data augmentation technique. Regarding the recall, accuracy, f1 score, and precision. Numerous experiments' findings show that CNN outperforms DT, RF, and GBT. The accuracy score they found from CNN is

6 88.83%. which is far better than any other model.

This [18] article's author introduces a revolutionary technique that distinguishes between the skin disorders known as eczema hand, eczema subacute, lichen simplex, stasis dermatitis, and ulcers. They applied Derm-NN, a convolutional neural network application to identify skin illness, to image processing. In this study, a classifier prototype was created that, after assessing a picture and matching it to its prior training data, would identify the kind of skin illness. They utilized the Dermnet dataset for this test, and they randomly selected some photographs from the internet. The classifier comprises five categories for skin illnesses and can properly diagnose 70% of skin conditions. The picture data may be lost due to flaws in the digitization process. Picture enhancement techniques can be used to fix abandoned picture data. Before using the preparation approach of their own created CNN architecture, they performed an informational collection assortment, information resizes, information planning, and information expansion. They have 4 convolutional layers and 13 more layers in their CNN model. They obtained the images from Dermnet and also amassed images from the internet. Five different diseases have downloaded more than 500 images. The data set is balanced. Utilizing 5 different strategies, they have employed a data augmentation strategy to prevent overfitting of the data. 1. Make a 90-degree turn 2. Turn the dial 90 degrees. 3. Shading 4. A salt-and-pepper noise 5. Flip Horizontal using 100 x 100-pixel pixels. Their total image count after growing is 3000, of which 2400 were used for testing and 600 for preparation. Whereas the F1-score was 2.45, their overall accuracy was 0.73.

In this paper[11], the author introduces a unique technique for categorizing skin lesions into one of five groups. The processing of images is accomplished by the utilization of a hybrid method, which may include deep CNN and SVM with error-correcting output codes (ECOC). To extract the features, the pre-trained CNN model known as "AlexNET" was utilized, and the ECOC SVM classifier, which reduces the multi-class classification issue to a set of binary classification problems, was used to classify the data. 15,546 photos that may be used for educational reasons for free from sites like DermIS, DermQuest, and DermNZ were used in the experiments. The writers have compiled a list of healthy categories. The training and testing datasets for the data store's photographs are split 70:30 apart. Preprocessing and 227x227 resizing are performed on the photos. The method of cross-validation using 10 folds, which employs the 10th subset for testing purposes and the other nine for training and calculating the average error across all 10 trials, was employed to prevent overfitting. The SVM classifier achieved an accuracy rate of 86.21% on average. The precision of the intelligent expert system that was suggested earlier has increased by 3.21% over the previous work.

In [23], the author used 4 distinct methodologies using deep learning algorithms to distinguish between skin diseases such as Cellulitis Bacteria, Impetigo Bacteria, Ringworm Fungus, and Sporotrichosis Fungus. The dataset consisted of 97 images of infected skin and healthy skin. The dataset was preprocessed using Image Resizing, Label Encoding, Image Augmentation, and other image preprocessing techniques to better suit their model. All input images are resized into 256x256 to be forwarded to the convolution layer. The 5 class labels were used, one for healthy skin and four for

7 various skin diseases, which were labeled using one hot encoding. The dataset was split into 80-20 measures, where training was done by 80%, other 20% was utilized in the evaluation of the training model. A fine-tuned version of the VGG16 model was used with 150 epoch cycles. Their implemented model achieved a test accuracy of 86%.

In [6], the author focuses on utilizing image processing techniques and Artificial Neural Networks (ANNs) to detect dermatological diseases. 9 types of diseases were taken to classify between them. The aim is to automate the process of disease identification and improve diagnostic accuracy in dermatology. The dataset taken was a combination of images of the infection as well as user input data. A total of 775 pictures were taken in the image dataset. Through the use of pre-processing techniques, two distinct types of characteristics were retrieved and they are YCbCr, Grey Scale image conversion, sharpening filter, 5*5 matrix Median filter, Smooth Filter, and Binary Mask. the dataset was split 10-15-75 for Both semi-supervised and supervised ANN algorithms were used. Among the diseases, foot ulcer and vitiligo were detected with the highest accuracy at 97% and acne was detected with the lowest amount at 85%.

The paper [17] centers around the implementation of cutting-edge deep-learning strategies and the classification of medical data to forecast instances of bacterial skin infections. The objective of the author was to demonstrate that deep learning models, such as CNN, are capable of automating and improving the accuracy of forecasting skin illnesses that are caused by bacterial agents. The data here was to the author ran ResNet50, VGG16, InceptionV3, VGG19, and Xception architectures and compared the results. The dataset used was collected from the public domain Kaggle. The total number of images used was 3300, with a split at 80% train and 20% test split. The image was preprocessed with 224x224 resizing and RGB conversion and normalization. Signmoid and ReLU were used as activation functions. After applying the algorithms, the author found that VGG16 provided 91.303% accuracy, the highest among all the models.

The study [24] aimed to test 2 things, the capability of the computer-aided pre-screening process of erythema migrans how much better it is compared to regular humans, and how the model works on different datasets. The authors reference a survey that has shown that the respondents correctly identified EM 73% of the time that had central clearing which is easier to detect. The detection rate drops to 30% when affected with vesicular, uniformly red, or disseminated manifestations. Due to the lack of any dataset with proper labeling, they decided to collect some EM images from the internet and also collect some from isolated clinical environments. They also added HZ and tinea corporis as confuser pathologies. This allowed them to do a 4-way classification and also a binary one. Their reason for doing a test on a 4-way classifier is because even though TC and HZ might look similar to EM their treatments are way different. They split the data into 70%, 10%, and 20% for train validation and testing respectively. Their chosen model was ResNet50. Of the total 1834 image datasets, the first test of 4-way classification resulted in 84%, and binary had 90.20% accuracy, which is a promising score for a pre-screener. The second experiment which was to test the clinical image with training from the online images also proved to have great results with 86.53% accuracy. This means that the pictures don't need to be taken in a particular isolated environment to be detected accurately. The author talks about the limitations of their study that they had to parse out inappropriate or irrelevant images which made the dataset smaller 8 and made it easier to identify. Another limitation that was mentioned was due to the lack of online data on darker-skinned individuals. They concluded that they demonstrate how DCNN can be promising to automate the pre-screening process of EM. The computer performs better than individuals with no or some training at all times except one individual. Deployment of such a technique could prevent many misdiagnoses and stop many long-term complications associated with Lyme disease.

This paper [29]. is about the development and evaluation of a new image-explain ability method called ensemble AI explainability (XAI) for a deep learning model. This method combines SHAP and Grad-Cam++ techniques. In this paper, it discussed the COVID-19 impact on global healthcare. Existing literature gaps in deep learning algorithms explaining ability in images are discussed here. The contributions include proposing ensemble XAI and creating an evaluation checklist. LIME is introduced here as well. It is an interpretable model approximation method. It is very time-consuming.

The research paper titled 'Explainable Artificial Intelligence (XAI) for Deep Learning Based Medical Imaging Classification' by the authors Rawan Ghnemat, Sawsan Alodibat and Qasem Abu Al-Haijapropose [25]. propose a methodology and the experimental findings for COVID-19 detection from chest X-ray (CXR) images via deep learning algorithms, as well as the use of explainable AI (XAI) for model interpretability. The study emphasizes the importance of improving model interpretability through explainable AI to get a better grasp of the method of choice that the model uses in medical imaging applications. The study makes use of five different CXR image datasets for binary (normal vs. COVID-19) and multiclass (COVID-19, pneumonia, and normal) classification. The paper highlights the use of a local interpretable model (LIME) for generating heat maps to visualize the important regions of the CXR images that have contributed to the classification decision. Different evaluation metrics such as accuracy, precision, recall, loss, and area under the curve (AUC) are examined for assessing the model's performance. Their findings include that VGG16 with the LIME model generates the most accurate results while using fewer parameters. The results of the experiment demonstrate the model's ability of accurately identify and differentiate between COVID-19, pneumonia, and normal cases from the CXR images. The author talks about some limitations of their study such as the model being incompatible with the numerical dataset as it was meant for image data, the data was taken from patients in the early stage of COVID-19 which may leave out many other variants or late stages for the AI to accurately predict. The authors also propose that this process can be applied to CT scans in the future. Overall, the research aims to provide insights into how a prediction is made and build trust in machine learning systems instead of blindly trusting a black box for medical applications by leveraging visualization and explainable models.

2.2 Background

2.2.1 Explainable AI - LIME

LIME is a common widely used XAI technique. Most of the deep neural network models are black box models, which generate decision functions that can not be interpreted by humans easily. The purpose of LIME is to explain the predictions of these black-box models by generating locally faithful and interpretable models. It operates under the assumption that a complex model's behavior can be approximated accurately in a local region by a simpler, interpretable model. LIME starts by perturbing the input data instances around the instance of interest. This involves creating slightly modified versions of the original data samples. The perturbed data instances are then fed into the black-box model, and the corresponding predictions are recorded. The equation of the lime is as such:

$$\xi(x) = argmin_{g \in G} L(f, g, \pi_x) + \Omega(g) \tag{2.1}$$

After that, in the perturbed space, the new locally trained model is implemented. As the model is working on a single data, it can be summed in a local classifier equation, through which interpretation is possible. The coefficients or rules of this interpretable model are used to provide insights into the factors influencing the model's prediction. LIME can be applied to any deep learning model without knowledge of its internal structure.

Chapter 3

Methodology

In this section, a clear description of the approach taken for this research will be described. The step-by-step description includes the collection and structure of the dataset, the preprocessing method, the system architecture, the description of the classifier models, and the description of the XAI models.

3.1 Dataset Collection

Skin rash is caused by multiple factors hence. In this research, we selected three distinct rashes, named drug rash, pityriasis rosea, and ringworm rash as classes for our collection of distinct images of skin diseases, the data had to be classified and labeled accordingly for the model to accurately process the data.

In terms of skin diseases, labeled authentic image data has been scarce to come by that can be reliably used in the detection and classification. Although numerous datasets related to skin diseases exist, the lack of labeling in such datasets proved to be a significant hurdle in skin disease detection-related research. To address this challenge, we gathered our dataset from a publicly available resource namely Kaggle and DermNet.

As we had multiple sets of datasets, we decided to merge both datasets to make a unified dataset with more instances. The merging of datasets enabled the models to train and test on more instances, making sure the classifier interpretation was more accurate as the classifier learns patterns, trends, and correlations from more instances.

3.2 Dataset Description

The merged dataset comprises a collection of 1,381 distinct images portraying various skin diseases. The dataset was precisely categorized into three distinct classes, namely drug rash, pityriasis rosea, and ringworm rash, encapsulating the specific types of skin conditions.

The images in this dataset were meticulously collected through web scraping, ensuring a robust and authentic representation of the three distinct rash classes. A manual filtration process was implemented to guarantee the reliability and authenticity of the images selected for inclusion in the dataset.

In organizing the dataset for model training and evaluation, we adopted an 80-20 split. 80% of the dataset, 1111 images, was allocated for a training set. The remaining 20% of the dataset, 270 images, was reserved for the testing set. The drug rash, pityriasis rosea, and ringworm rash classes in the training dataset consist of 426,467 and 218 images respectively. All images are in JPG(Joint Photographic Experts Group) format.

3.3 Instances of The Classes in The Skin Disease Dataset

Here, 3.1 shows instances of pictures of the three classes of rashes that are present in the skin rash dataset and used to implement the models mentioned in this paper.

Figure 3.1: Example data of 3 classes(Pityriasis rosea, Drug rash, and Ringworm) from the dataset we implemented

3.4 Data Preprocessing

Data pre-processing is the preliminary step that diverts the raw data gathered from sources into more relevant information to achieve consistency in values and information. Before utilizing any models, the dataset would be preprocessed to maximize authentic outcomes from the model. A pre-processed dataset ensures minimum levels of redundancy and inaccuracies in values, thus helping the model to produce maximum accuracy.

As we are using CNN-based architectures, the model requires images to be of the same dimensions, which ensures the filters extract meaningful features to predict the results. Here for this dataset, the images are resized as 224x224x3.

Here, below Table 3.1, we have shown the image augmentation parameters. After resizing the dataset, the training dataset was put through a dataset augmentation procedure. Data augmentation is a technique used in deep learning to artificially increase the size of a training dataset by applying various transformations to the existing data. This increases the dataset by creating new variations of the original data without changing its underlying features, thus increasing the datasets that the classifiers can work on. This procedure allowed the expansion of the training

Table 3.1: Image Augmentation Parameters

Parameter	Values
rescale	1./255
rotation range	20
width shift range	0.2
height shift range	0.2
shear range	0.2
zoom range	0.2
horizontal flip	True
fill mode	nearest

dataset, thus eliminating any chance of undertraining the dataset and ensuring the reduction of over-fitting, thus improving the model's overall performance.

3.5 System Architecture for Predicting A Skin Disease by Utilizing Deep Learning Models

In this research, the idea is to make a working model that classifies skin rashes. In Figure 3.2, the steps of the model have been discussed. The model extracts the necessary features from the image data and predicts the class of the image according to the learned data and characteristics.

For this purpose, various deep learning architectures will be used, more specifically, deep neural networks will be utilized for their advantage in handling complex and non-linear relationships between input and output and parallel processing. The Deep Neural Network algorithms selected from this model are MobileNetV2, ResNet50V2, DenseNet201, and Convolution Neural Network. The current models using deep learning classifiers use small datasets along with small levels of accuracy in predicting these skin diseases accurately. After that based on the result analysis, the best classification algorithm would be determined.

Our model implementation would try to improve on this aspect by implementing more datasets as well as more models to choose the model that best fits the data from this particular category. To improve the accuracy of the training, the dataset has been augmented in the preprocessing step. After classification, the result would be compared using different techniques, and the feature detection technique would be compared.

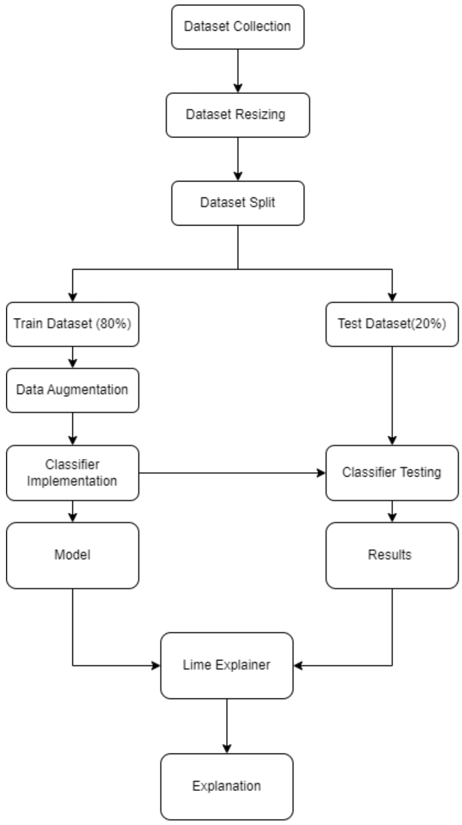

Figure 3.2: Process Flow for Classifying Between Skin Rash classes

Implemented Models:

3.6.1 MobileNetV2

MobileNet is a lightweight Deep CNN model, that implements depthwise separable convolution and also gives highly accurate results even in less memory applications [8]. It is a smaller and more mobile architecture based on the CNN as mentioned above architecture. The Depthwise separable convolution filters used in MobileNet are composed of two filters. First, the depthwise convolution filter gives a single convolution for every input channel it connects with [13]. First, the layer gives a 1*1 convolution with ReLU. The other layer takes the result to depthwise convolution and the third mixes it with 1*1 convolutions. The other filter takes the result of depthwise convolution and mixes it with 1*1 convolutions. The input image measures 224 x 224 x 3 pixels in size.

$$\hat{\boldsymbol{G}}_{k,l,m} = \sum_{i,j} \hat{\boldsymbol{K}}_{i,j,m} \cdot \mathbf{F}_{k+i-1,l+j-1,m} \tag{3.1}$$

The primary focus for MobileNets is working on latency and creating smaller-sized architecture [8]. MobileNet models are considered extremely important in cases where the model is implemented on mobile and embedded devices. The small size of the architecture enables these models to be run on lower-powered processing devices. For embedded and mobile devices, the model is based on a simplified design that assembles consolidated deep neural networks of low latency using depth-wise separable convolutions [8]. In MobileNet, number of parameters is significantly lower than in a network with conventional convolutions of a similar depth, like CNN. This results in deep neural networks that are compact. Figure 3.3 is the structural view of the mobilenet architecture.

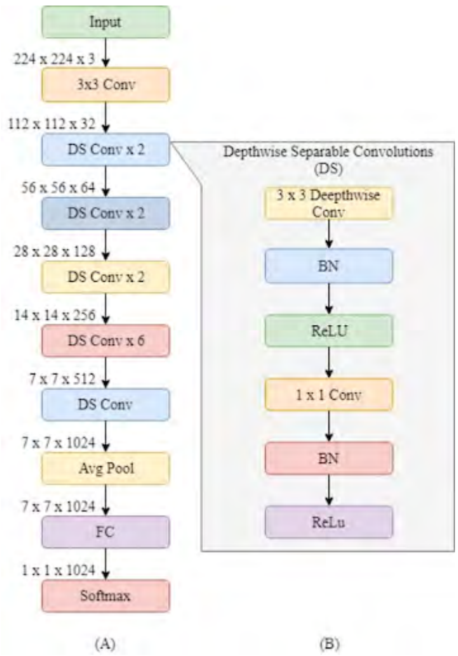

Figure 3.3: Structure of MobileNet Architecture

3.6.2 ResNet50V2

ResNet is a widely used model, that is based on the CNN architecture. Resnet50 is a variant of ResNet architecture, that has 50 total layers, 48 of which are Convolution Layers, 1 MaxPool layer, and 1 Average Pool Layer. It uses forward propagation and Backpropagation to generate the prediction values. It solves the problem faced by VGG16, which is stacking more layers one over another. As models stack one layer over another, a highly packed model with lots of layers generates inaccurate data. As network input goes from one layer to another in linear form between the convolutional layers[7]. This stacking of layers creates the opportunity for the low-value layer to be created which hampers accuracy. The ResNet Model Bypasses that by utilizing a connection called a Residual Network skip connection, which creates a path from the convolution layers to bypass low-value data layers, thus effectively stacking layers without losing any accuracy. The residual layer is implemented to boost the overall accuracy of the architecture. If we take a subnetwork of certain stacked layers and if its performance function is H(x), then the residual function would be:

$$F(x) = h(x) - x \tag{3.2}$$

And the output of the subnetwork would be:

$$y = F(x) + x \tag{3.3}$$

As the model also backpropagates, the outcome function for that is similar to the subnetwork forward propagation result function. The photo input size for the network is 299 x 299.[1]. Here below, Figure 3.4 is the structural view of the ResNetV2 architecture.

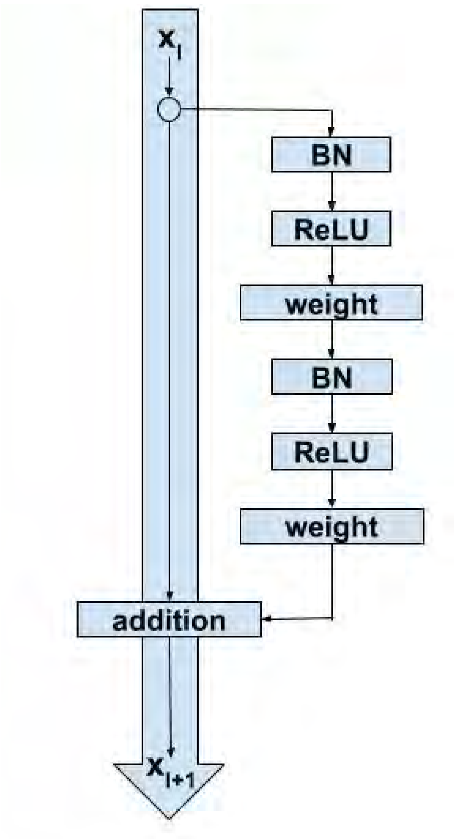

Figure 3.4: Structure of ResNetV2 Architecture

3.6.3 DenseNet201

DenseNet201 is a deep convolutional neural network architecture that belongs to the family of Densely Connected Convolutional Networks or DenseNets. The key factor for this model is the densely connected layers, which are different from the Convolution Neural Networks or ResNet. In DenseNet, each layer of the network receives input from all the preceding layers, which is a key innovation from CNN. Also, by concatenation, DenseNet can merge the layer output with skip connections, which differentiates it from the ResNet architecture.DenseNet201 specifically has a deep architecture with 201 layers. It is designed to capture intricate patterns and hierarchical features in large-scale image datasets. The architecture includes densely connected dense blocks, which consist of multiple convolutional layers, and transition layers that control the spatial dimensions and the number of channels. Each of the layers in the DenseNet201 Model implements a non-linear transformation of

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \tag{3.4}$$

can be any composite function of operations such as Batch Normalization (BN), rectified linear units (ReLU), Pooling, or Convolution (Conv)[9]. The dense connectivity in DenseNet helps address the vanishing gradient problem and facilitates feature propagation throughout the network. This results in better model performance, improved training convergence, and efficient parameter usage. DenseNet201, being a deeper variant, is more suitable for complex tasks and datasets with a large number of classes.

3.6.4 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a widely used Deep learning architecture, that was modeled inspired by the vision of living beings [10]. CNN detects images by applying them through a 3-layer architecture, which is convolutional, pooling, and fully connected layers[10]. The convolutional layer's goal is to learn input feature representations. A feature map's neurons are linked to an area of nearby neurons in the preceding layer. The equation of feature value is calculated by:

$$z^l_{i,j,k} = \mathbf{w}^l_k{}^T \mathbf{x}^l_{i,j} + b^l_k \tag{3.5}$$

CNN gets an advantage in reducing the complexity of the model using the weight-sharing mechanism. Nonlinearities are introduced into CNN through the activation function, which is beneficial for multi-layer networks that detect nonlinear attributes. The commonly used activation functions, in this case, are Sigmoid, tanh, and ReLU[10]. In this research, the ReLU activation function has been used to activate corresponding neurons. VGG16 is used as the base model which is a pre trained model based on imagenet. Then comes Flatten which transforms the output from the previous layer into a 1D array, preparing it for the fully connected layers followed by Dense layer and BatchNormalization layer. Again, we used a dropout layers which helps prevent overfitting. One or multiple fully connected hidden layers are implemented at the end of the model that make the decisions and predictions. [10]. Below, Figure 3.5 shows the structural view of the CNN Architecture.

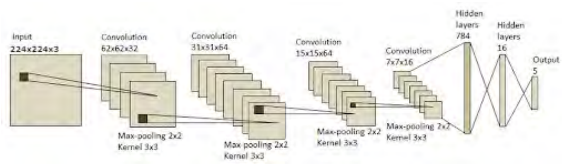

Figure 3.5: Structure of CNN Architecture

CNN achieves high accuracy on big datasets via the join feature and classifier learning, putting it an advantage over other architectures [10]. CNN is also a great model for image datasets because of its minimal preprocessing requirements for original images [5], and it can train large datasets with high accuracy by implementing millions of parameters [14].

Chapter 4

Implementation & Result Analysis

Our implementation technique is to implement the aforementioned 4 algorithms in the preprocessed dataset and so analyze the model's effectiveness. For all the models, the prediction is done using different values. Our accuracy measurement depends on four parameters evaluated for the three classes, which are precision, recall, f1-score, and support score respectively.

For uneven datasets like ours, confusion matrices produce useful solutions. The performance of a classification model (or "classifier") on a set of test data for which the true values are known is often described by a confusion matrix, which is a table.

For this research on the detection of skin rashes, we ran all of our classifiers on virtual machines using Google Colab. Google Colab is a hosted Jupyter Notebook service, that can be used on a cloud-hosted CPU or GPU. The Colab runs on Ubuntu 64 bits and is composed of a single core hyper threaded Intel Xeon processor 2.3 GHz and 13 GB of RAM.

4.1 Result Analysis of Skin Rash Prediction Model

As we utilized 4 different architectures in our model, all the architectures were implemented sequentially. First, the CNN algorithm was implemented on the preprocessed dataset to check the viability of implementing CNN architecture for this dataset. To check the compatibility and results we use Train Accuracy vs. Validation Accuracy Graph.

In the context of training models like neural networks, the "train accuracy vs. validation accuracy" graph is a visual representation that shows how a model's accuracy varies while being trained on both the training dataset and a different validation dataset. The model's training accuracy reveals how accurate it is concerning the training data. The validation accuracy curve displays the model's performance on a validation dataset that is separate from the training dataset, demonstrating its capacity to predict fresh data.

As we are using both heavy and mobile models, we run the risk of overtraining light models if we select a high epoch number, and undertraining heavier models if we train a small number epoch. So we varied the epoch number based on the model. The batch size is always kept at 32 regardless of the model.

Here for the Fig 4.1 CNN algorithm, we see that, after 30 epochs the training accuracy is 91%, signaling high levels of accuracy in training. Also, we see a significant improvement in the level of validation accuracy as the epoch increases, which stays around 90% at 30 epochs in which the CNN model was trained without any overfitting or underfitting.

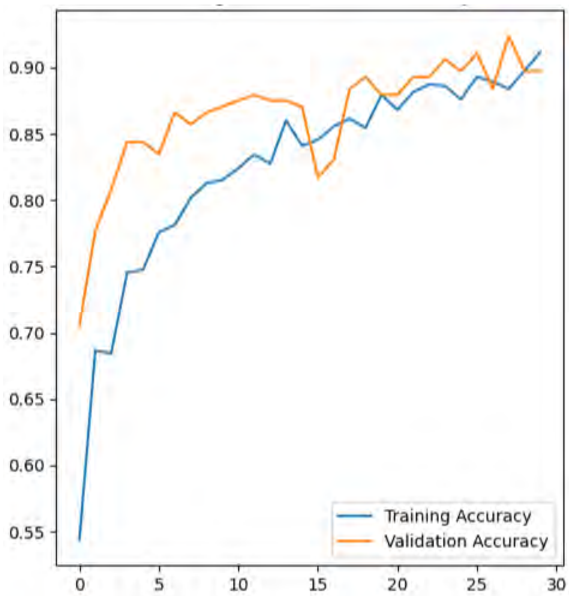

Figure 4.1: Training vs Validation Accuracy CNN

Next, we implemented the MobileNet Model Fig 4.2. As it is a light mobile model, the convolution layers trained much faster than the other models due to the smaller number of layers, hence the model was trained for only 10 epochs. Which worked very well for the overall dataset. The accuracies after implementing the dataset remained constant and gradually increased after implementing all epochs. The Train Accuracy for the MobileNet was 80% and validation accuracy was 83%, suggesting a good fit of the dataset.

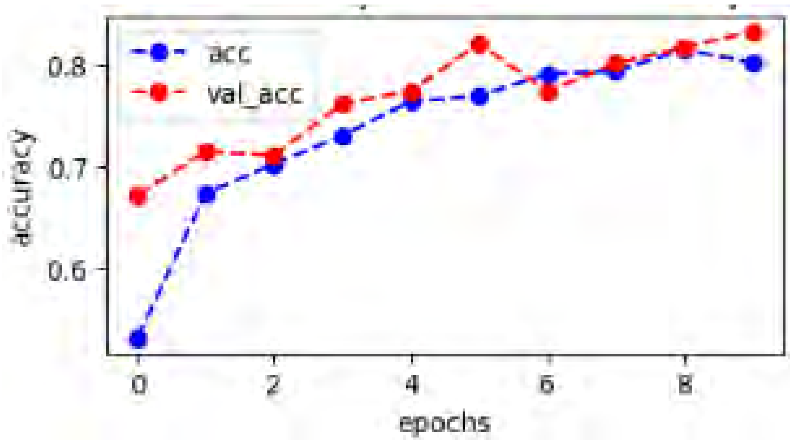

Figure 4.2: Training vs Validation Accuracy MobileNetV2

Here for the Fig 4.3 DenseNet201 algorithm, we ran it for 30 epochs. We found the training accuracy at 87%, signaling high levels of accuracy in training. Also, we see a significant improvement in the level of validation accuracy as the epoch increases, which stays around 87% at 30 epochs in which the DenseNet201 model was trained with a perfect fit

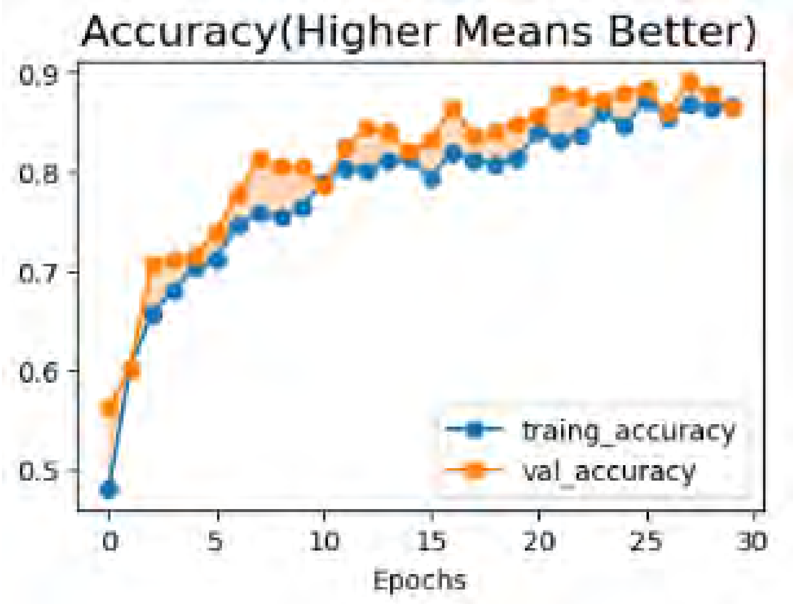

Figure 4.3: Training vs Validation Accuracy DenseNet201

Here for the Fig 4.4 ResNet50v2 algorithm, we ran it for 30 epochs. We found the best training accuracy at 94%, signaling high levels of accuracy in training. Also, we see a significant improvement in the level of validation accuracy as the epoch increases, which stays around 93% at 30 epochs in which the ResNet50v2 model was trained with a perfect fit.

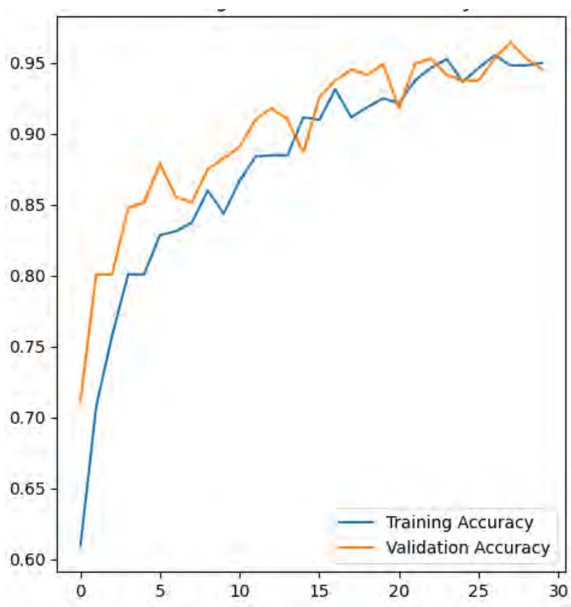

Figure 4.4: Training vs Validation Accuracy ResNet50V2

4.1.1 Comparison of All Models Scores

Here in Table 4.1, we compared the different scores of the neural network classifiers we used in the model. In this research, we used the accuracy, precision, recall, and F1 score to describe the model's performance. The accuracy score in machine learning is a statistic that measures the percentage of accurate predictions made by a model out of all predictions made. Accuracy is a widely used metric but it is not a very good metric if the dataset is imbalanced or skewed towards a specific class. Hence we used recall, precision, and F1 score. Recall is a statistic used to evaluate a classification model's effectiveness, particularly in binary classification problems. It evaluates the model's propensity to locate each pertinent example (true positive) inside a class. On the other hand, the precision score shows the ratio of true positives to the total of true positives and false positives. The F1 score combines both precision and recall to create a single number that represents the model's accuracy.

Table 4.1: Performance Matrices for Severity Detection

Model	Precision	Recall	F1-score	Accuracy
CNN	0.82	0.81	0.81	0.81
MobileNet	0.91	0.90	0.90	0.90
ResNet50V2	0.92	0.91	0.91	0.94
DenseNet201	0.88	0.87	0.87	0.87

In summary, after implementing our model, the ResNet50V2 model gave out the highest percentage of accuracy with 94%. It was also the most precise in prediction with a precision score of 92% and a recall score of 90% showing that it got the classes correct. ResNet50V2 was closely followed by MobileNetV2 with 90% accuracy, 91% precision, and 90% recall, and 90% F1 score. DenseNet201 and CNN gave decent numbers of accuracy, 87% and 81% respectively. From table 4.1, we decide that Resnet50V2 and MobileNetV2 are greatly accurate with this dataset. As the dataset is fitted with multiple algorithms, the dataset is very enriched and of high quality despite being imbalanced in one class.

4.1.2 Confusion Matrix Explanation

The confusion matrix enables us to understand the predictions by the trained classifiers we implemented in the model. It helps to identify the best working model by explaining the precision and recall value generated, by every class. It is a matrix with true labels compared with predicted labels. This enables us to identify the classifier's lackings on the dataset due to its architecture or the lackings of the dataset itself. As our dataset is an imbalanced one, this enables us to identify where the problem arose due to this imbalance in every classifier.

Here, from the confusion matrix, we can identify that all the models work very well while running the model. This backs up the classification table 4.2 Resnet50V2 Being the better working model is accurately identifying all the classes. Followed by the MobileNetV2 model and DenseNet201 model which show a decent performance in predicting classes.CNN model has an underwhelming performance here.

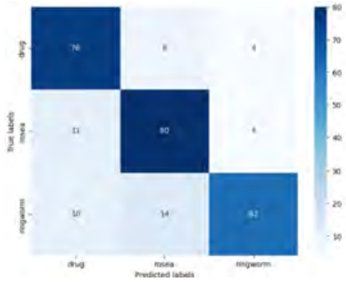

Figure 4.5: CNN

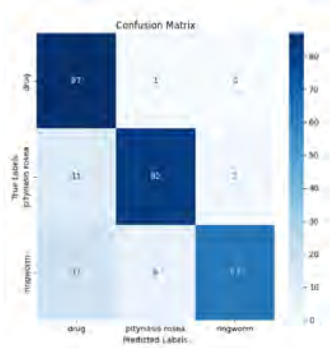

Figure 4.6: MobileNet

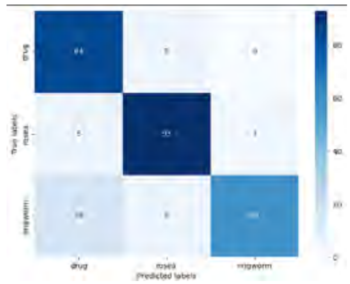

Figure 4.7: DenseNet201

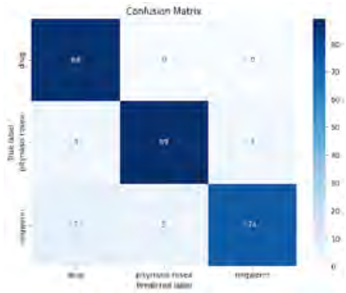

Figure 4.8: Resnet50V2

From the matrices, we have identified an anomaly in that ringworm has an underwhelming performance, it is being predicted significantly less compared to other classes, as we have less data in ringworm due to an imbalanced dataset, it is the notion that the training of the ringworm class has been insufficient compared to the other classes.
Also, another anomaly we see is that the majority of the wrong predictions we are having are being predicted as drug rash, regardless of the original class. Hence, there could be an issue in the feature selection of the models.

4.2 XAI Explanation

This section of the paper will explain Deep Learning algorithm using LIME. LIME XAI utilizes interpretable machine-learning techniques to demystify complex AI models by highlighting influential features and explaining predictions for local data. Here, for image datasets LIME explains the classifier regions for selecting features in a single instance otherwise known as the local instance. Here, to explain the feature selection of every classifier, we selected 3 distinguished pictures of every class and ran LIME on that data for every classifier. Figures 4.9, 4.10, and 4.11, show the pictures in a tabular format. These pictures have been used to implement LIME to identify the features used for predicting the models.

Figure 4.9: Drugrash.

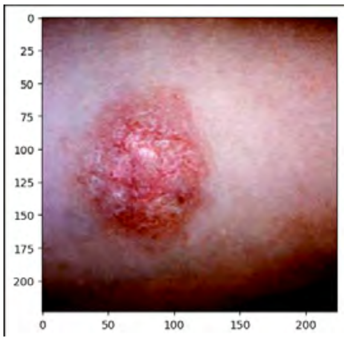

Figure 4.10: Pityriasis Rosea

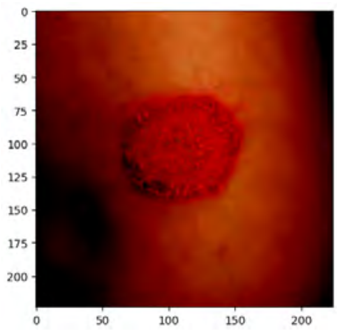

Figure 4.11: Ringworm

4.2.1 LIME Explanation of The Feature Selection

In Figures 4.12, 4.13, 4.14, 4.15 we have compiled all the LIME implementations that have been implemented on the models. Here we can see all the features that have been taken into consideration during the prediction. We see, that all the models have considered the right features of the images for the predictions. Resnet50V2 has worked quite well in recall and precision. From the figures above we can see that both the classifiers have considered background. However, the pre-trained ResNet50V2 could detect enough of the disease feature to get the right prediction. The pretraining of the model can be a factor in this case.

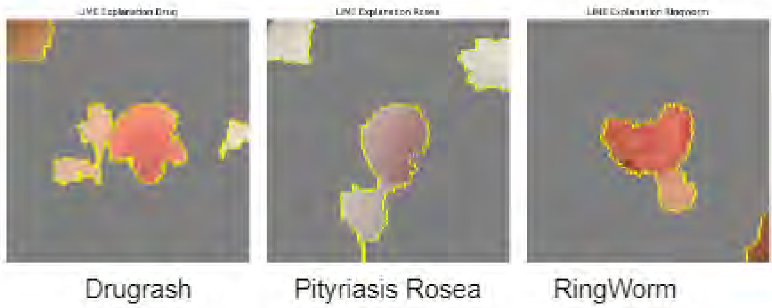

Figure 4.12: CNN

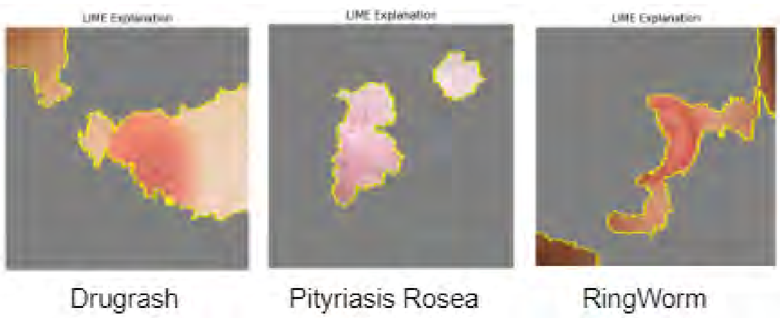

Figure 4.13: MobileNetV2

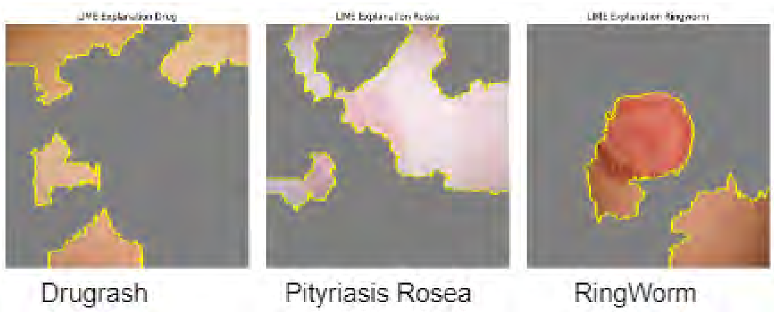

Figure 4.14: ResNet50V2

We see an anomaly in CNN too as it does detect all the disease-ridden areas for all classes but the recall and precision score is not as good as we see from This Local Explanation. DenseNet also shows a similar characteristic.

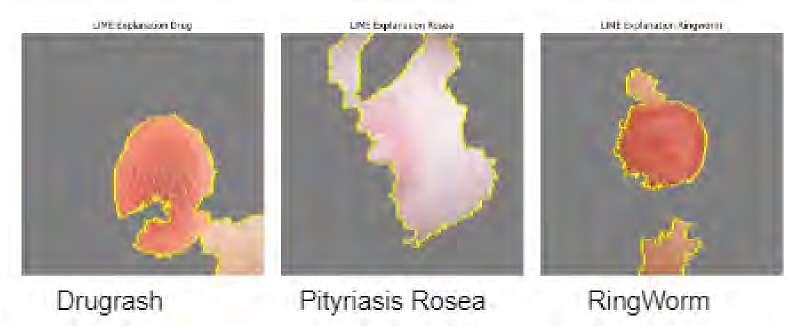

Figure 4.15: DenseNet201

4.2.2 Lime Explanations of The Anomalies

As we have seen in the confusion matrix table, some of the Pityriasis Rosea and Drugrash, especially in DenseNet201 and CNN. To explain this, we have taken the CNN model and seen the comparison between different images and their predictions by the models. The class we have implemented this finding is Pityriasis Rosea.

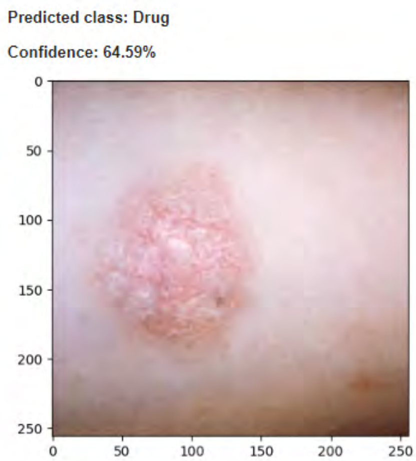

Figure 4.16: Prediction

In Figure 4.16, Pityriasis Rosea has been detected as a Drug Rash, But in Figure 4.17, the data have been accurately identified.

From the lime explanation both in the previous figures 4.12 to 4.15 we see that the incorrectly predicted data has similarities with the Drugrash, as it is a close-up picture of the disease. As we see the CNN model focuses on the rashes only, the model is predicting wrong due to the similarities. But in the whole body image, the data is being predicted right due to taking other parts as features as well. We saw a training class of the dataset in Drugrash that had a lot of images in closed-up orientation. Hence, we can say that the Closed images are being wrongly predicted by the CNN and DenseNet model due to the features it is taking in. ResNet50v2 and MobilenetV2 avoid this by taking in other models' features as well.

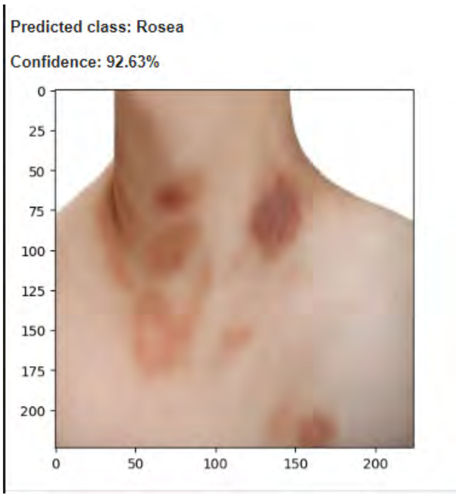

Figure 4.17: Pityriasis Rosea prediction

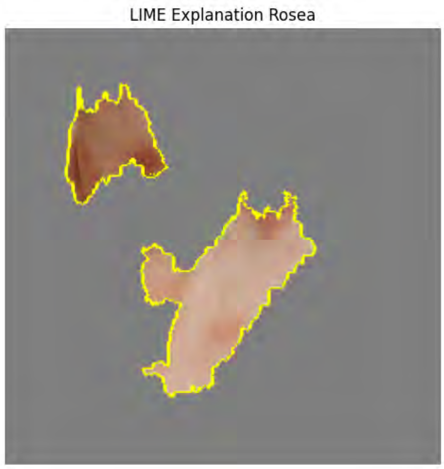

Figure 4.18: LIME explanation

4.3 Findings

This entire research can be yielded into the following valuable findings:

- The Accuracy of all the classifiers in the dataset is above 80%.This indicates the high quality of the overall dataset despite being imbalanced.
- Augmentation of the dataset, helps small datasets thus can perform well even in heavier models.
- ResNet50V2 and MobileNetv2 perform the best in the model. On the other hand, CNN's performance was comparatively very low. This indicates the usage of the pre-trained model can be very beneficial due to already being adapted to other images.
- The success of ResNet50v2 and MobileNetV2 can be attributed to their capability of taking into consideration the features that are nearby the wound, as only considering the wound can sometimes result in improper assessment, which happened in the case of CNN and DenseNet201.

Chapter 5

Conclusion & Future Work

5.1 Conclusion

In this paper, we presented the applications of Deep Neural Network Models in the detection of Pityriasis rosea, Drug rash, and Ringworm skin rash. We collected image data of these classes from publically available sources and used 4 different models, CNN, MobileNetV2, ResNet50 V2, And DenseNet201. To better the results we augmented the dataset. From the acquired results after implementing the models, we see all the models give good accuracy, all the models have a great precision, recall, and F1 score, signifying this dataset is a good one and the models we chose are accurate in this research. ResNet50V2 and MobileNetV2 work the best with this type of data, with ResNet50V2 achieving the highest accuracy rate of 94% and MobileNetV2 having an Accuracy rate of 90%. This result implies that the dataset acquired is of high quality and works for multiple architectures. Through the implementation of lime, we saw that the models are accurately using the features and the local bias of the dataset does hamper a bit of the accuracy for CNN and DenseNetV2 model.

5.2 Limitations and Future Work

It is very challenging to implement Deep Learning algorithms in the healthcare field for it's limited availability of datasets. We faced quite a lot of challenges in this regard due to not finding a large dataset that is classified properly. The large dataset can improve the accuracy of the hindered models. Also, the LIME model is only a local interpreter of the feature a model implements to get an even better result. We should Implement any global interpreter of Explainable AI to get a picture of the feature importance for the whole dataset. For this SHAP is a good global Explainer Model.

Despite having some limitations this thesis performed all the predictions and provides prominent future works. It is our belief in the future with the right tools and circumstances we can have even better outcomes for the model.

Bibliography

[1] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[2] S. Pillai, C. Oresajo, and J. Hayward, "Ultraviolet radiation and skin aging: Roles of reactive oxygen species, inflammation and protease activation, and strategies for prevention of inflammation-induced matrix degradation–a review," *International journal of Cosmetic Science*, vol. 27, no. 1, pp. 17–34, 2005.

[3] A. Morrone, "Poverty, dignity, and forgotten skin care: Dermatology in the stream of human mobile population," *Dermatologic clinics*, vol. 26, no. 2, pp. 245–256, 2008.

[4] N. Ahmed, M. Islam, and S. Farjana, "Pattern of skin diseases: Experience from a rural community of bangladesh," *Bangladesh Medical Journal*, vol. 41, no. 3, pp. 50–52, 2012.

[5] X.-X. Niu and C. Y. Suen, "A novel hybrid cnn–svm classifier for recognizing handwritten digits," *Pattern Recognition*, vol. 45, no. 4, pp. 1318–1325, 2012.

[6] R. Yasir, M. A. Rahman, and N. Ahmed, "Dermatological disease detection using image processing and artificial neural network," in *8th International Conference on Electrical and Computer Engineering*, IEEE, 2014, pp. 687–690.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[8] A. G. Howard, M. Zhu, B. Chen, *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[9] G. Huang, ""densely connected convolutional networks."," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[10] J. Gu, Z. Wang, J. Kuen, *et al.*, "Recent advances in convolutional neural networks," *Pattern recognition*, vol. 77, pp. 354–377, 2018.

[11] N. Hameed, A. M. Shabut, and M. A. Hossain, "Multi-class skin diseases classification using deep convolutional neural network and support vector machine," in *2018 12th International Conference on Software, Knowledge, Information Management & Applications (SKIMA)*, IEEE, 2018, pp. 1–7.

[12] N. S. A. ALEnezi, "A method of skin disease detection using image processing and machine learning," *Procedia Computer Science*, vol. 163, no. 4, pp. 85–92, 2019.

[13] R. Poojary and A. Pai, "Comparative study of model optimization techniques in fine-tuned cnn models," *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA) IEEE*, 2019.

[14] R. Poojary and A. Pai, "Comparative study of model optimization techniques in fine-tuned cnn models," in *2019 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, IEEE, 2019, pp. 1–4.

[15] N. Barr Kumarakulasinghe, ""evaluating local interpretable model-agnostic explanations on clinical machine learning classification models."," *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS),*, 2020.

[16] F. A. Mphande, "Skin disorders in vulnerable populations: Causes, impacts and challenges.," *Springer*, 2020.

[17] S. A. Nayyef, L. M. Hasan, and A. A. Ibrahim, "Using deep learning and medical data classification for predicting bacterial skin infection," in *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, IEEE, 2020, pp. 1–9.

[18] T. A. Rimi, N. Sultana, and M. F. A. Foysal, "Derm-nn: Skin diseases detection using convolutional neural network," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, IEEE, 2020, pp. 1205–1209.

[19] S. A. AlDera and M. T. B. Othman, "A model for classification and diagnosis of skin disease using machine learning and image processing techniques," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, 2022.

[20] T. H. Aldhyani, A. Verma, M. H. Al-Adhaileh, and D. Koundal, "Multi-class skin lesion classification using a lightweight dynamic kernel deep-learning-based convolutional neural network," *Diagnostics*, vol. 12, no. 6, p. 2048, 2022.

[21] V. R. Allugunti, "A machine learning model for skin disease classification using convolution neural network," *International Journal of Computing, Programming and Database Management*, vol. 3, no. 7, pp. 141–147, 2022. DOI: 10.33545/27076636.2022.v3.i1b.53.

[22] 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*, vol. 12, no. 9, p. 426, 2022.

[23] V. R. Pai, S. G. Pai, P. Suhasi, and P. Rekha, "Identification and classification of skin diseases using deep learning techniques," 2023.

[24] P. M. Burlina, ""automated detection of erythema migrans and other confounding skin lesions via deep learning."," *Computers in Biology and Medicine, vol. 105,*, Feb. 2019.

[25] R. Ghnemat, ""explainable artificial intelligence (xai) for deep learning based medical imaging classification."," *Journal of Imaging, vol. 9, no.*, 9, Sept. 2023.

[26] S. Shi, "A modified perturbed sampling method for local interpretable model-agnostic explanation.," *arXiv:2002.07434, arXiv,*, 18 Feb. 2020.

[27] M. R. Zafar and N. M. Khan, "Dlime: A deterministic local interpretable model-agnostic explanations approach for computer-aided diagnosis systems," *arXiv:1906.10263, arXiv,arXiv.org,*, 24 June 2019.

[28] Zhang, "Explainable ai in deep reinforcement learning models for power system emergency control."," *EEE Transactions on Computational Social Systems, vol. 9, no. 2*, Apr. 2022,

[29] L. Zou, ""ensemble image explainable ai (xai) algorithm for severe community-acquired pneumonia and covid-19 respiratory infections."," *IEEE Transactions on Artificial Intelligence, vol. 4, no.*, 2, Apr. 2023.