Skin Disease Detection and Classification Using Deep Learning

by

Mehedi Hasan Shuvon
18101686
Rowshanara Sadia
18101188
Shanjida Habib Shormi
18101097
Umma Tania Arafin
18201203
Md. Rawha Mikdad Chowdhury
18101672

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 2022

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:

মেহেদী হাসান শোভন

Mehedi Hasan Shuvon

18101686

Rowshanara Sadia

Rowshanara Sadia

18101188

Shanjida Habib Shormi

18101097

Umma Tania Arafin

Umma Tania Arafin

18201203

Md. Rawha Mikdad Chowdhury

18101672

Approval

Skin Disease Detection and Classification using Deep Learning" submitted by

1. Mehedi Hasan Shuvon (18101686)
2. Rowshanara Sadia (18101188)
3. Shanjida Habib Shormi (18101097)
4. Umma Tania Arafin (18201203)
5. Md. Rawha Mikdad Chowdhury (18101672)

Of Fall, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science and Engineering on January, 2022.

Examining Committee:

Supervisor:(Member)

Md. Khalilur Rhaman

Md. Khalilur Rhaman, PhD

Associate Professor
Department of Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, PhD

Designation
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Skin Diseases have been the primary focus of this study, as they are one of the most lethal diseases if not diagnosed and treated early. The research will enable the fields of Medical Science and Computer Science to collaborate in order to save lives. Although Machine Learning, Deep Learning, and Image Processing have been utilized previously to treat skin diseases, we are attempting to improve the accuracy of this work by implementing new models of Image Processing and Deep Learning. The purpose of this research is to demonstrate how to accurately diagnose Skin diseases at an early stage using the optimum model. Here we have used three distinct neural models to classify a custom dataset. We further analyzed the accuracy of the MobileNetV2, InceptionV3, and ResNetV2 to come up with an optimized model that can be configured further to a mobile application for vast use. We built the architecture on more than 1450 images representing nine distinct skin disorders in comparison with fresh skin. We carefully compared our data and classified it based on the images of our customized dataset. Finally, we determined the nine diseases with a 96.77% accuracy with the help of MobileNetV2 which is our ideal model for the goal we want to achieve.

Keywords: Image Processing; Machine Learning; Deep Learning; MobileNetV2; InceptionV3; ResNetV2; Epoch; Softmax; Skin Disease; KNN; Classification; CNN; Detection; Accuracy; Validation; TensorFlow; Keras; Keras Layer; Dense Layer

Acknowledgement

First and foremost, we give thanks to the Almighty Allah for allowing us to finish our thesis without any serious setbacks. Second, we would like to express our gratitude to our adviser, Md. Khalilur Rhaman, PhD, Associate Professor, BRAC University, sir, for his valuable assistance and guidance. He was always willing to assist us whenever we needed it. Finally, we want to thank the judging panel for all of their constructive criticism, which has really aided us in refining our work. Finally, without our parents' unwavering support, we might not be able to achieve our goals. We would not be on the verge of graduating if it weren't for their kind support and prayers. All of the individuals named here deserve our gratitude.

Table of Contents

List of Figures

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

ACC Accuracy

$ADPI$ Advanced Configuration and Power Interface

BCC Basal Cell Carcinoma

CAM Class Activation Map

CNN Convolutional Neural Network

$FFNN$ Feedforward Neural Network

FPR False-Positive Rate

$ICECA$ Initiative for Cultural Exchange and Computer Arts

$IEEE$ Institute of Electrical and Electronics Engineers

$ISIC$ International Skin Imaging Collaboration

KNN K-Nearest Neighbors

MAP Mean Average Precision

ROI Region of Interest

SK Seborrheic Keratosis

SUT Suranaree University of Technology

TNR True Negative Rate

TPR True Positive Rate

Chapter 1

Introduction

The skin is the body's primary organ, and it serves a variety of tasks, including protection against physical, chemical, and biological threats. Even though skin disorders are widespread among many poor nations, they have not been recognized as a severe problem that may benefit from public health interventions. In poor countries, skin disorders are one of the most common health issues. Environmental, economic, literacy, ethnic, and social norms all have an impact on skin disorders. Here, Bangladesh is a densely populated, developing nation with humid weather. Thus, skin diseases are quite frequent among the population. Patients do not seek therapy until pushed by the severity of their symptoms due to a lack of information. Because of its undeveloped economy and social backwardness, Bangladesh has a high rate of infectious diseases. A significant prevalence of cutaneous infection has also been seen in the rural community. Bangladesh has about 140 million people living in a 55,598 square mile area with a population density of 2,639 persons per square mile, according to the World Bank Development Indicators Database (April 2006). In Bangladesh, nearly 19% of all OPD patients have minor to severe skin issues. Acne, Tinea, Basal Cell Carcinoma, Actinic Keratosis, Viral Skin Infections, Bacterial Skin Infections, and Deep Dermal Problem are the most frequent illnesses observed by dermatologists and researchers in Bangladesh. Because rural people visit the doctor less frequently, illnesses quickly pose a life-threatening threat to them. The majority of skin disorders are, once again, infectious. As a result, they have an impact not just on themselves but also on people who live near them, such as family members. This problem brings us to work with technologies for early-stage skin disease classification and detection. Thus, our thesis is based on skin disease classification and detection where we are using a comparison between MobileNetV2, ResNetV2 and InceptionV3 which are commonly used neural network architectures for Image Detection and Classification. Among them, the architecture of MobileNetV2 is based on an inverted residual (IR) structure. Compared to the other models, which use extended representations for input and output, the residual block's input and output are tiny bottleneck layers. The intermediate expansion layer incorporates MobileNetV2 filters using lightweight depthwise convolutions. Furthermore, we uncover how crucial it is to reduce non-linearities in the thin layers in order to maintain representational power. By demonstrating how this improves performance and describing the rationale for this strategy. Ultimately, this technique automatically disconnects the input/output domains from the expressiveness of the transformation, laying the groundwork for more investigation.

Our performance is evaluated against ResNetV2 and InceptionV3. We examine the swap between accuracy and the number of multiply-adds (MADD, leading) operations and the number of parameters also. Therefore, we got 96.77% accuracy with a validation of 97.50% accuracy using MobileNetV2.

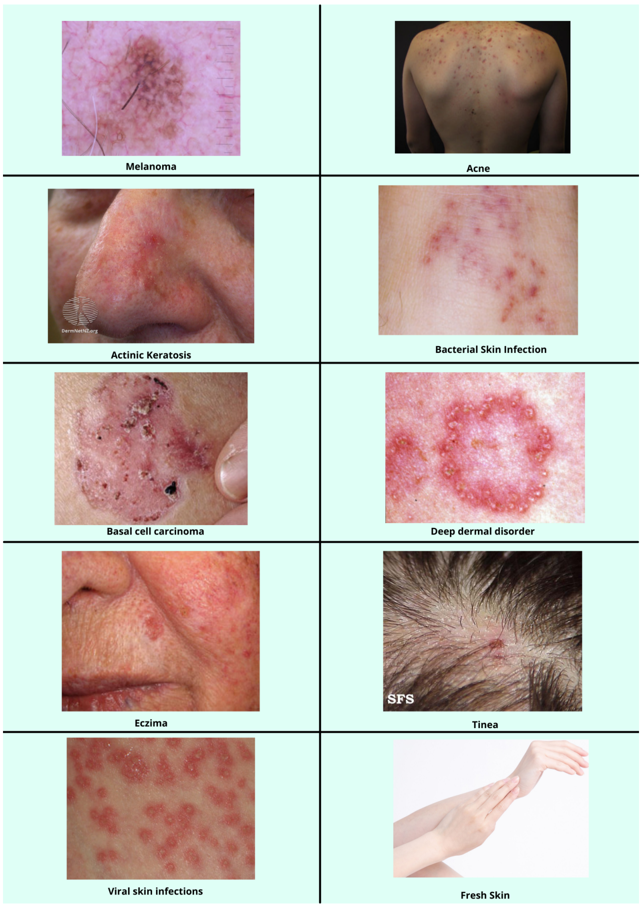

Figure 1.1: Nine different skin diseases(Acne, Basal cell carcinoma, Actinic keratosis,Viral Skin infections, Bacterial skin infections, Deep dermal disorder, Tinea, Eczema)

1.1 Research problem

Although doctors are the best way to identify and cure a disease, the improvement of technologies made us aware of the implementation of Deep Learning into the medical field. Thus different kinds of models can be used for image classification but selecting and implementing the most suitable algorithm for our model was a tough job. Previously research papers working in the same field either had an insufficient dataset or a lower level of accuracy. The suitable models were also missing and a comparison of different models to get the optimum result was not seen in the previous works. Thus it required more research to at least prevail the path for future works. Hence this research did not only come up with the optimum model, customize large dataset but also implemented the Unsupervised Approach.
Image processing from a specific image and extracting each pixel to match with a particular dataset using deep learning and finding proper labels is the most challenging task for the whole project. However, without an adequate dataset that includes necessary titles, any activity can not find the affected area from an image. Again, the validation of a dataset is significant. A dataset without proper authentication may mislead the whole model to a completely different result. Therefore, selecting an authentic, valid and trustworthy dataset is the most crucial part of this model. Data bias in machine learning refers to an error where some sort of databases are more likely to happen and may cause a significant problem in the entire model, leading the whole model to failure. When a dataset includes a specific number of data many times, the machine takes the data and gets biased on those numbers responsible for showing the result using repeated data [2]. Data repetition makes a high bias in a dataset that is accountable for overfitting the model. However, an overfit model is an example of a complicated model that makes the whole model inefficient and ineffective [16]. The very challenging and demanding job is to understand the entire medical terminologies about skin diseases to label out in the dataset. Again, some other kinds of languages are generally unknown, making issues to understand and work accordingly. For detecting and classifying skin diseases by image processing and comparing it with the dataset, we will need deep learning, which can be solved using two kinds of approaches. They are:-

1. Supervised Approach
2. Unsupervised Approach

1.1.1 Supervised Approach and Unsupervised Approach

Different kinds of algorithms like K-NN, Naive Bayes, or SVM supervised approach are being used. Selecting and implementing the most suitable algorithm for our model is a tough job. Algorithms are the way of representing the whole model. Therefore, to maintain the proper process of finding the affected cell and work accordingly and train the model fast and efficiently, we have to select and focus on the most suitable algorithm or modified version[3]. Image processing is the process by which after collecting the samples from the cell to detect diseases from the given image by breaking it into pixel by pixel is hard work to do. While processing the image, the quality, consistency, and stability of the light source need to be considered. Again, unenhanced or low-quality pictures will not break down properly, which will be responsible for a poor model and trained dataset. Therefore, the whole activity should be done carefully with proper concerns and precautions while doing image processing. Getting image data with proper quality is not always possible. Thus image processing enhancement is always needed before using the raw data. Image enhancement refers to the process of highlighting particular information of an image and formatting all unnecessary information as the specification needs. For example, noise reduction, revealing blurred details, and genuine level adjusting to underline important features of an image. This research will use various image enhancement approaches to find the best solution to tackle poor-quality ideas in the dataset.

1.2 Research objectives

With the assistance of deep learning and image processing, this study attempts to identify and categorize nine distinct skin disorders (Melanoma, Acne, Basal cell carcinoma, Actinic keratosis, Viral Skin Infections, Bacterial Skin Infections, Deep dermal disorder, Tinea, Eczema) more correctly and quickly. The study also successfully classifies the fresh skin to identify the diseases more accurately. The purpose of implementing the Fresh skin dataset is to be more focused on the diseases so that the program can identify if a skin has any disease or it is normal without any illness. Image processing identifies exact usable information, and deep learning learns and extracts valuable data from the raw data and classifies automatically. The following are the research goals:

1. Profoundly recognize affected skin cells and the process of causing damage.
2. Classify the differences between diseased and non-diseased skin.
3. Profoundly acknowledging given nine diseases and making detection approaches, like deep learning as well as image processing.
4. Presenting the classification and probability for each disease as percentage, where the highest percentage holder disease will be under consideration of being affected with.
5. Enhancing a model for detecting and classifying diseases using Deep Learning and Image Processing.
6. Evaluating the process and algorithm of the model.
7. Providing necessary proposals on enhancing the model.
8. Processing the model in such a way that it can identify only the affected cells taking into consideration the skin color tones.
9. Providing a comparison between models to get optimum results.

Chapter 2

Detailed Literature Review and Related Work

The performance of six proposed techniques (Balanced Random Forest, Balanced Bagging, AdaBoost, Random Forest, Logistic Regression, Balanced Bagging & SVM) was evaluated using 2,453 photos. Dermoscopic photos from seven skin illnesses (melanoma, melanocytosis, basal cell carcinoma) were utilized to create the model. In a recent study, Van-Dung Hoang, Chi-Mai Luong, Antoine Doucet, Cong-Thanh Tran & Tri-Cong Pham stated that the EfficientNetB4-CLF model had the greatest accuracy (89.97%), lowest recall (86.13%), and fewest recalls (0.39%). Using a minority dataset proportion until the proportion equals the maximum classes were introduced by them in data level techniques. DenseNet169 classified pictures better than EfficientNetB4. ImageNet's trained network provides the CNNs. To avoid overfitting in Custom Fully Connected Layers(CFCL), every concealed layer comes after a 0.2 dropout block. Their study employed Adam Optimizer to enhance their network as follows: amsgrad= Not True, epsilon=None beta 2=0.999, Ir is dynamic between 0.000001 & 0.00005. and CyclicLR to alter the LR one epoch at a time. Back-propagation is a technique used to increase the weight values of a CNN architecture. The authors used 24,530 images in this study. These pictures have been resized to 256*192px. The photographs are classified into test, train & validation where Test was 10%, Train was 80% & validation was 10%.[13].

Also, Neha Agrawal & Dr Sagaya Aurelia published a report on improving the detection of three diseases (Vascular Tumors, Vitiligo & Melanoma) utilizing transfer learning to strengthen the medical sector. They employed a model named Inception V3 which is a part of transfer learning. If we use a huge dataset it improves accuracy. At the time of writing, they used Kaggle data and found datasets of skin infection where accessible was for 117 types. Actually, major models were built in Keras, and the accuracy and loss were graphed using NumPy and Matplotlib. The learning rate of the Adam optimizer was 0.001. Here Used 21,808,931 trainable parameters and 34,432 non-trainable parameters. The Inception V3 model was trained with ten layers and sixteen clusters in 27 batches across ten epochs. The Adam optimizer used categorical cross-entropy as a loss function. Pre-training accuracy was 81.70 % and testing accuracy was 84.84 %, but fine-tuning increased detection accuracy by 12.79 %. [14].

A report named A Visually Interpretable Deep Learning Framework for Histopathological Image-Based Skin Cancer Diagnosis by some members of IEEE has explained some extents of a new architecture name DRANet. On the basis of a genuine histology image acquired over the last decade, the DRANet deep learning system proposes to distinguish eleven different forms of skin diseases. They propose the Class Activation Map (CAM) approach for visualizing deep neural networks. DRANet's tiny model size and high classification accuracy make it very practical. The model has an overall accuracy of 86.8%, a weighted average exactness of 88.8%, call back of 86.8% & an F1 score of 87.1% for eleven epidermis conditions. On average, the performance of RANet and DRANet-BN is enhanced by 3.8%, 5.9%, 3.8%, and 4.4%, respectively. DRANet outperforms EfficientNet-B0, ShuffleNet V2, MobileNet, NASNet Mobile & MobileNetV2(86.8% vs. 75.6 %, 78.2 %, 81.6 %, 77.8 %, 78.6 %). DRANet outperforms EfficientNetB1 on all measures except F1 score (86.8% vs. 86.5%) and recall (86.8% vs. 86.3%). (4084K vs 6589K). With 86.8% accuracy, DRANet is equivalent to InceptionV3 (86.3%)& ResNet50 (85.5%) which is a bit finer than 82.1 % of VGG19 & 82.1 % of VGG16 [15] .

Zhang, X., Wang, S., Liu, J. et al. Towards improving diagnosis of skin diseases by merging deep neural networks & knowledge of human beings, a report has shown deep learning algorithms were utilized in the study to aid in the diagnosis of recurrent affected skin disease predicted on dermoscopy images. In a test dataset of 1067 photos, their system had an accuracy of 87.25 2.24%. Seborrheic Keratosis (SK), Melanocytic nevus, Psoriasis & Basal Cell Carcinoma (BCC) were chosen during a previous study. They used the code package of GoogleNet InceptionV3, which had been pre-trained on approximately more than 1 million photos, to create their method. To find the inaccuracies, they employed subcategories inside categories. Furthermore, their results have a standard deviation of roughly 2–5%, which can reflect variance in precision. The vision is to build a decision support system that will aid clinicians in making more accurate diagnostic decisions, as well as a patient-friendly system accessible via mobile applications.[7]

Albawi, Saad et al. published Robust skin disorders detection and classification using deep neural networks. The ISIC dataset was used for Melanoma, Nevus, and Atypical Identification. Statistical region merging, iterative region merging, and adaptive threshold are also thought to be used for ROI extraction. The approach combines Gaussian and average filters to reliably forecast skin diseases. The author created an algorithm using a unique efficient ROI to detect skin lesions and it must be extracted before the feature extraction and classification processes. The authors created adaptive region expansion to improve segmentation. Convolutional neural networks can classify atypical, nevus, and malignant lesions (CNN). The method is tested in MATLAB's image processing application. The data shown are BCC, Dermatofibroma, LNOS, and Lichenoid Keratosis. It is split 80/20 between normal and abnormal images. Their study found the KNN classifier to be inefficient. Using a CNN classifier improves identification precision and specificity. Except for deep learning using CNN, FFNN surpassed all others. A total recognition rate is compared in their investigation. Their study focuses on Melanoma, Nevus, and Atypical and the algorithm worked well on the ISIC study dataset.[8].

Classification of Skin Disease from Skin Images Using Transfer Learning Technique was published in 2020 at the 4th ICECA. Recently, CNN's were utilized to detect skin cancer lesions. They used the ICIS open dataset. The data collection includes almost ten thousand photos of Melanoma and benign illnesses. CNN model 2016 extracted one thousand features from a single image by Janoria, Minj, and Patre. Both the input and output images use the CNN model's VGG 16 architecture for feature extraction. A useful model that can extract one thousand features from a particular image. Moreover, K Nearest Neighbor, Linear Discriminate Decision Tree obtained 99.9% accuracy. However, they discovered that Ensemble learning can only gain 48.2%. The transfer learning model performs admirably across a wide variety of Convolution layers & classifier combinations. Classification accuracy of above 95% for both the decision tree & the K nearest neighbor filters. They note that both classifiers' ROC curves indicate a near-zero false prediction rate. Experiments show that some models, like K closest neighbors, are above 95% accurate in many iterations. However, complicated models such as supervised methods with enhanced trees perform poorly, with an accuracy of less than 50%. Classifiers are defined as nonlinear binary classifiers, rather than coarse multiclass models.[11].

The report is named "Dermatological Classification Using Deep Learning of Skin Image and Patient Background Knowledge" was written by K.Sriwong, S.Bunrit, K. Kerdprasop, and Nittaya Kerdprasop in December 2019. This paper was supported by the Suranaree University of Technology (SUT). They utilized image data and patient background knowledge in their analysis, as well as deep learning approaches such as Unsupervised Pretrained Networks, Recurrent Neural Networks, Recursive Neural Networks, and CNN (using three primary layers). In this article, an automatic approach is used to classify skin diseases using the CNN deep learning model. To improve CNN's classification performance, they include both visual data and patient-specific knowledge into the modeling process. The experimental results using a public dataset indicate that the CNN model can classify skin illnesses with an accuracy of 79.29 %, while their proposed strategy of including patient-specific background knowledge during the modeling phase can increase the accuracy to 80.39 %.[10]

The strategy of combining multiple deep learning models offers the best recognition impact. Dermoscopy is an elevated imaging modality (non - invasive method skin imaging) that allows for the observation of the anatomy of the skin at the intersection of the lower epidermis and the upper dermis. ISIC 2018 is the most widely used skin disease dataset. Popular CNN architectures that are used for image recognition using Deep learning are:-DensenNet, Inception, VGG, AlexNet & ResNet. The multimodal fusion method based on deep learning is the most suitable Model method. The accuracy (Acc), mean average precision (MAP), false-positive rate (FPR), true positive rate (TPR) & true negative rate (TNR) are used to evaluate deep learning in skin disease (TNR).[12]

Chapter 3

Proposed Model (Methodology)

The purpose of the proposed model aims to detect and classify nine different skin diseases(Melanoma, Acne, Basal cell carcinoma, Actinic keratosis, Viral Skin infections, Bacterial skin infections, Deep dermal disorder, Tinea, Eczema) more accurately and rapidly with the help of deep learning algorithm and image processing method. In this method, the model takes the picture of the affected area of the patient as input. After that, evaluates it methodically, and provides estimates of being plagued by the designated nine disorders in the anticipated proportion. Figure 3.1 depicts the model design at a high level.

Our proposed model consists of three major stages:

1. **Input data preprocessing:** This step is focused on structuring the input data so that our model can handle it easily.

2. **Processing:** This step involves utilizing IWC to organize input data into categories and creating a Decision Tree for categorization.

3. **Predictions:** This step is about making predictions using the Decision Tree that has previously been constructed. From then, the decision tree model will determine the extent to which the picture fits the database in terms of predicted percentage. IWC preprocesses the raw data before using it to build nine clustering, which together includes input from nine different skin issues. (Melanoma, Acne, Basal cell carcinoma, Actinic keratosis, Viral skin infections, Bacterial skin infections, Deep dermal problem, Tinea, and Eczema). Clustering is done based on the similarity of data characteristics. The preprocessed input data is divided into two groups after clustering: one for training and constructing the decision tree, and the other for evaluating the decision tree's accuracy in discriminating between all nine illnesses.

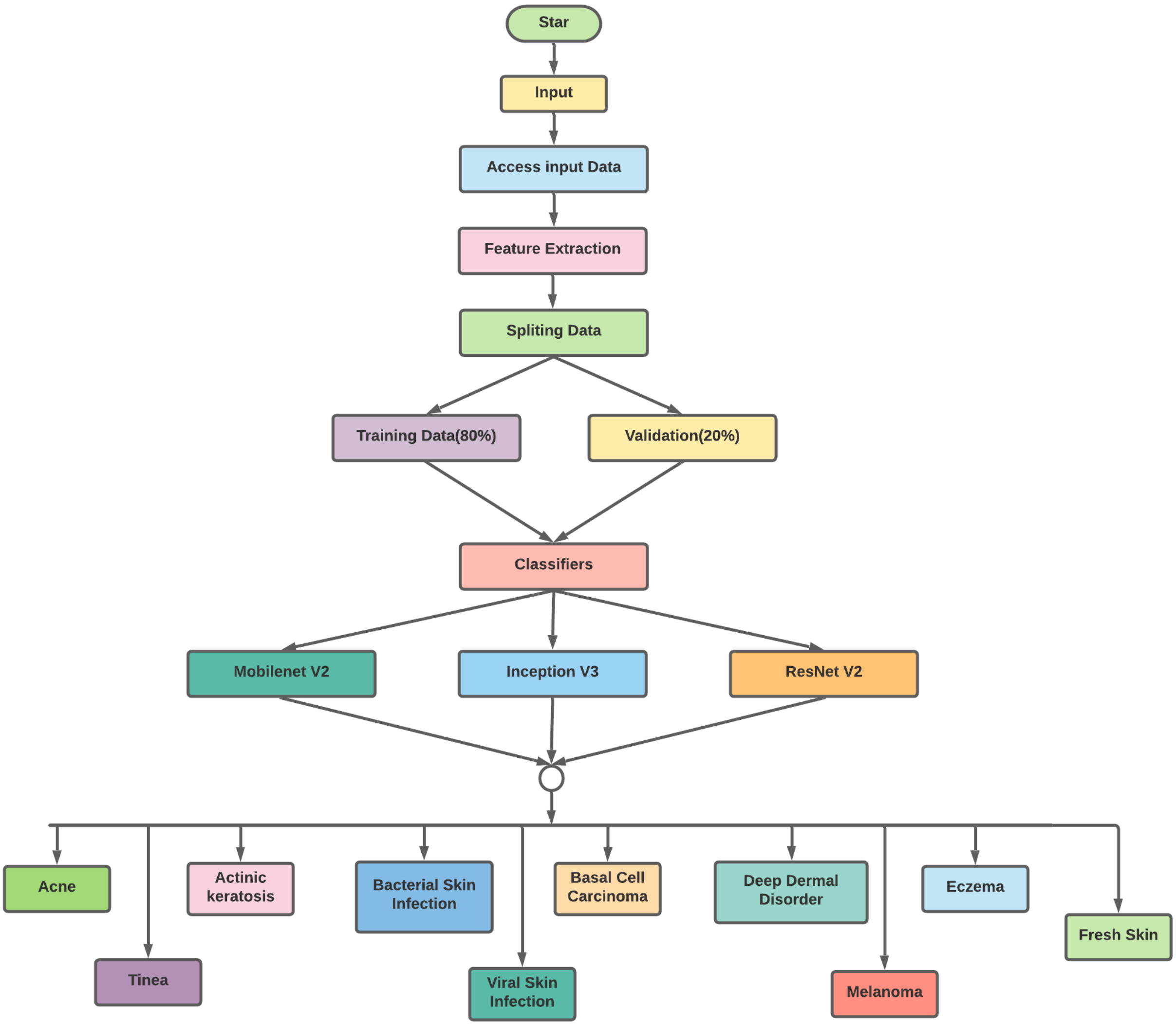

Figure 3.1: The flow chart of the proposed skin disease classification mode

3.1 Input data

Labeled data is very difficult to come by when it comes to input data for picture detection and classification. On the other hand, there are plenty of datasets available freely for particular skin diseases. However, it has been tough to label the dataset based on the diseases and skin color tone. Therefore, even if datasets like HAM 10000 [6] consist of thousands of records collected from different authentic sources and medical experts, we can not directly utilize them for the classification of various skin diseases considering our native skin tone as well. To overcome such challenges, an unlabeled and custom dataset was acquired. The dataset contains more than thousands of images for all the nine skin diseases collected from trusted medical web pages and databases like dermis.net[4], atlasdermatologico.com [3], etc.

The raw dataset downloadable link is given below:
https://drive.google.com/drive/folders/1yDVwvxbD4wlZpVo6VOn5mYdMPPsWREkr?usp=sharing

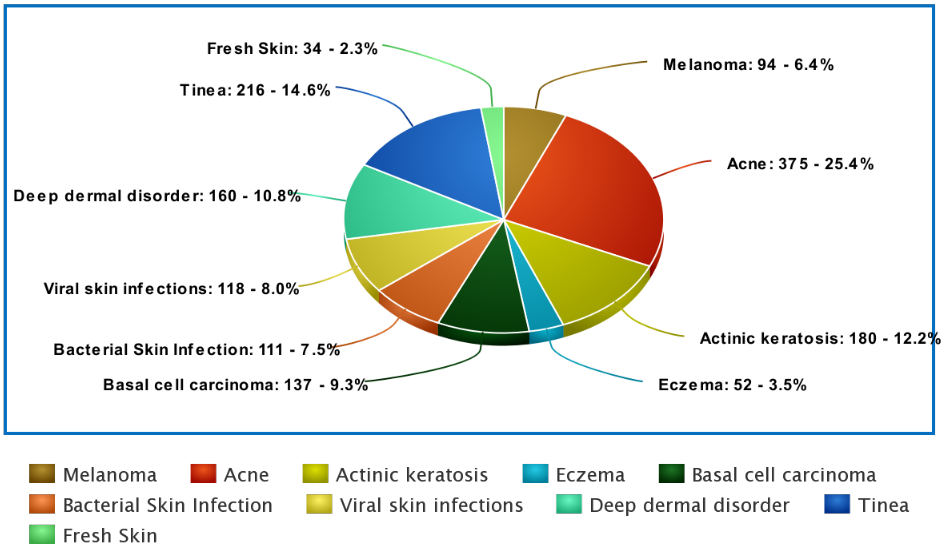

Figure 3.2: Pie Chart including the types of the diseases(Melanoma, Acne, Basal cell carcinoma, Actinic keratosis, Viral Skin infections, Bacterial skin infections, Deep dermal disorder, Tinea, Eczema)

Figure 3.2 shows each one type of diseases(Melanoma, Acne, Basal cell carcinoma, Actinic keratosis, Viral Skin infections, Bacterial skin infections, Deep dermal disorder, Tinea, Eczema) our custom dataset can detect and classify. As we built our custom dataset it is not labeled initially. Therefore, it requires processing to use it for the detection and classification of Skin diseases in this research.

3.2 Data Pre-processing

Processing large amounts of data with labeling can be computationally expensive. Therefore, a necessary amount of samples of readings was taken from various authoritative sources and placed in a .csv Microsoft Excel file named "Data.csv" which has mainly two columns for "image id" and "disease name". Moreover, any unnecessary attributes(i.e. columns) such as disease details and information about the patients, etc were deleted since our proposed model is only concerned with the detection and classification of skin diseases. In addition to this, rows were added to the input data manually (i.e. 1450+ records). The benefits of taking such an approach are:

1. The input data will be stored with proper labeling and will be easier to express.
2. The input dataset will become properly suitable for research.
3. Prevail over the issue of acquiring labeled dataset will lower the cost and computational expenditures

The resulting input dataset would be designed of two attributes (i.e. columns) which are id(image id) and Disease(disease name) accordingly. The value of each attribute in each record is actual as given in figure 3.3.

1	id	Disease
2	TSR_1_376	Acne
3	TSR_1_377	Acne
113	TSR_1_706	Actinic keratosis
114	TSR_1_707	Actinic keratosis
115	TSR_1_708	Actinic keratosis
116	TSR_1_709	Actinic keratosis
117	TSR_2_36	Eczema
118	TSR_2_37	Eczema
119	TSR_2_38	Eczema
120	TSR_2_39	Eczema
121	TSR_2_40	Eczema

Figure 3.3: Input data set with labeling before preprocessing

3.3 Deep Learning Approach

The design of MobileNetV2, a Convolutional Neural Network(CNN) which is mobile-friendly. A lightweight model MobileNetV2 which helps in image classification. It's built on the structure of inverted residual, with bottleneck levels connected by residual connections. In the intermediate expansion layer filters, lightweight depthwise convolutions are utilized as a source of non-linearity. Overall, MobileNetV2's architecture contains a fully convolutional layer with 32 filters, followed by 19 residual bottleneck layers. The MobileNetV2 architecture uses an inverted residual structure (MBConv Block CNN architecture), with small bottleneck layers at the input and output of the residual blocks. There are two main types of blocks in MobileNetV2. A 1-stride residual block is the first. Another one is blocked with a 2 stride for shrinking.

1. Both of the blocks have three levels.
2. 1x1 convolution with ReLU6 is the initial layer.
3. The depthwise convolution is the second layer.
4. In the third layer, another 1x1 convolution is employed, but this time there is no non-linearity. Deep networks will only have the power of a linear classifier on the non-zero volume part of the output domain if ReLU is applied again, according to the statement.

Input	Operator	Output
$h \times w \times k$	1x1 conv2d , ReLU6	$h \times w \times (tk)$
$h \times w \times tk$	3x3 dwise s=s, ReLU6	$\frac{h}{s} \times \frac{w}{s} \times (tk)$
$\frac{h}{s} \times \frac{w}{s} \times tk$	linear 1x1 conv2d	$\frac{h}{s} \times \frac{w}{s} \times k'$

There seems to be a part of growth. In all of the basic studies, t=6 was used. The actual result would've had 64xt=64x6=384 channels if the input had 64 channels.

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

Here; the expansion factor is T, the number of output channels is c, the repeating number is n, and the stride is s.

Chapter 4

Alternative Approaches (Comparison)

.We tried to build a comparison between several models to validate our claim, even though the provided model offers us the predicted output. Therefore we applied InceptionV3 and ResNetV2 in our dataset. The applications gave us some insightful predictions and values which helped us to validate our proposed model more. Both of the models are popular TensorFlow neural models for Image Classification applying Deep Learning approach.

4.1 InceptionV3

nceptionV3 is the Architecture of Convolutional Neural Networks which has been trained on the dataset of ImageNet with a good accuracy. Complex problems like classification problems are being solved using InceptionV3.In the ImageNet dataset, there are around 1000 classes and Millions of images. The Inception architecture is based on the idea of estimating and covering an ideal local sparse structure in a convolutional vision network with readily available dense components. Our network will be totally made up of convolutional construction blocks because we're assuming translation invariance. All we have to do now is find the best local structure and spatially duplicate it. It's a layer-by-layer method in which the correlation data from the final layer is examined and clustered into high-correlation groups of units. These clusters, which are linked to the units of the prior layer, constitute the units of the following layer. Each unit from the previous layer is assumed to correspond to a specific region of the input picture, and these units are organized into filter banks.

Correlated units would cluster in small areas in the lower layers (those closest to the input). This means that a significant number of clusters will be localized in a single region, which will be covered by an 11-layer convolutional layer in the next layer.

Again, fewer more spatially dispersed clusters will be covered by convolutions across larger patches, and the number of patches spanning larger and larger regions will decrease. In order to reduce patch alignment difficulties, current iterations of the Inception architecture are limited to filter sizes 1X1, 3X3, and 5X5, but this decision was chosen more for convenience than necessity.[1]

4.1.1 InceptionV3 Module

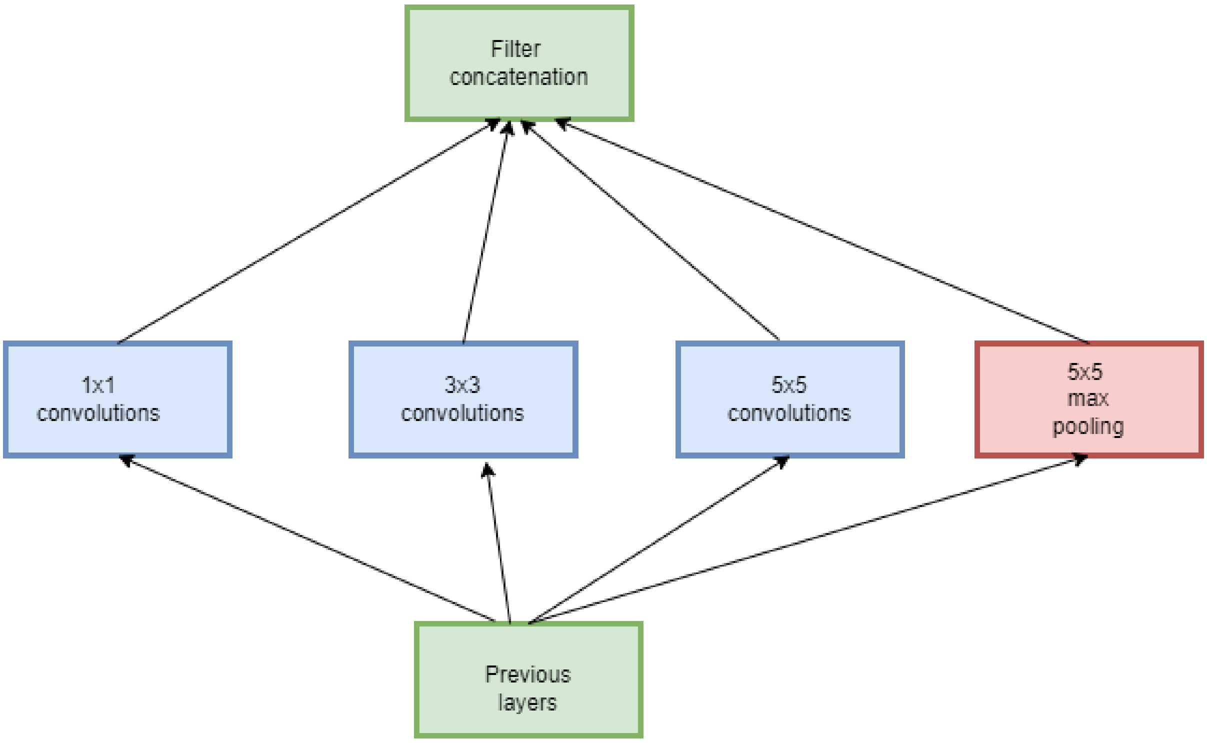

Figure 4.1: Traditional Networks

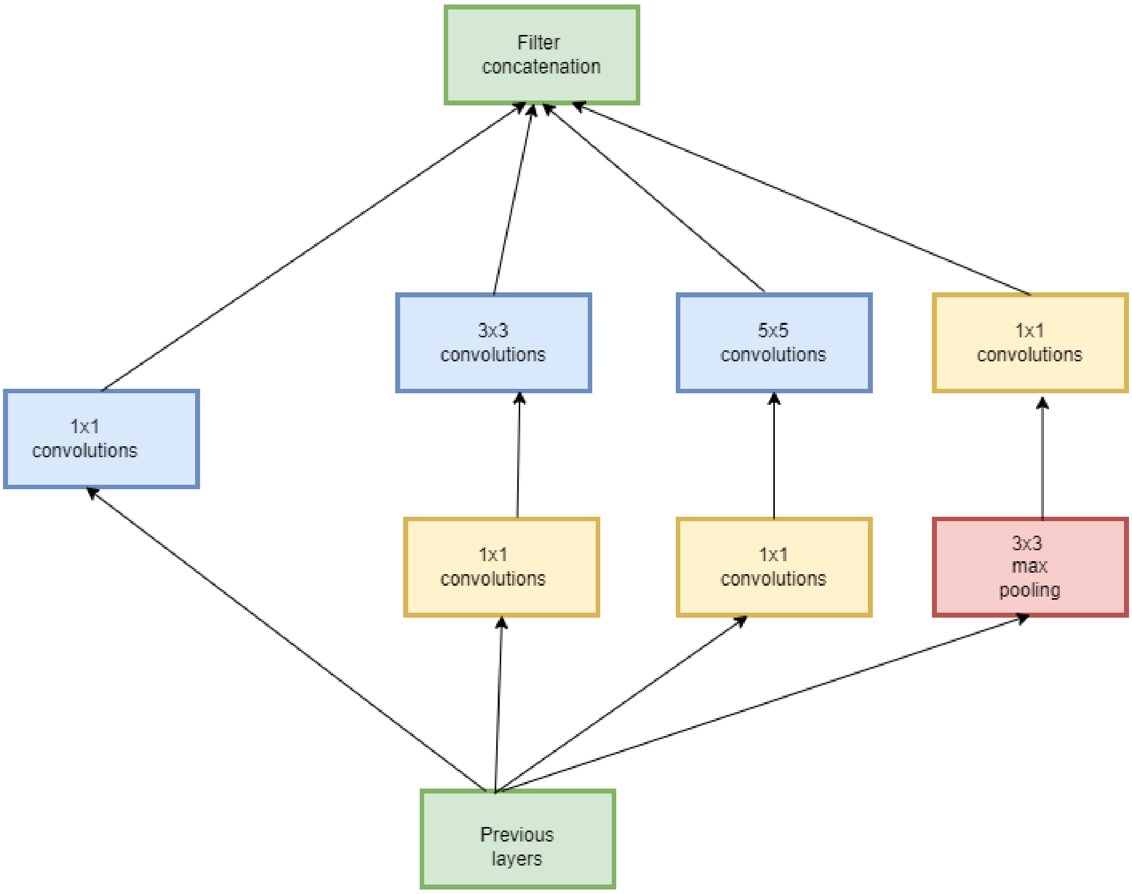

Figure 4.2: Inception Module with dimension reductions

4.2 ResNetV2

ResNet (short for Residual Networks) is a well-known neural network that is used to power numerous computer vision applications. This model took first place in the ImageNet challenge in 2015. ResNet50 consists of 48 Convolution layers, 1 MaxPool layer, and 1 Average Pool layer in a ResNet variation. ResNet became the most intriguing thing to happen in the computer vision and deep learning sectors after AlexNet won first place in the LSVRC2012 classification challenge in 2012.

In deep neural networks it's a major problem in vanishing gradient. Residual Networks can make the solution by vanishing gradient problem. Residual Networks use skip connections concept which is to add the original input to the output of the convolution block/layer.

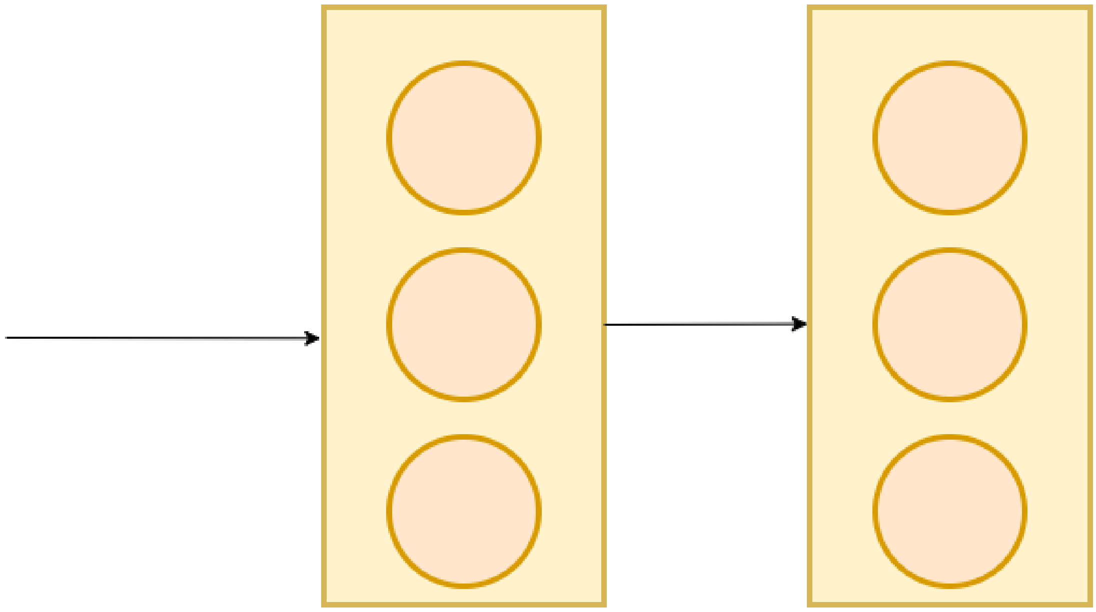

Figure 4.3: Traditional Networks

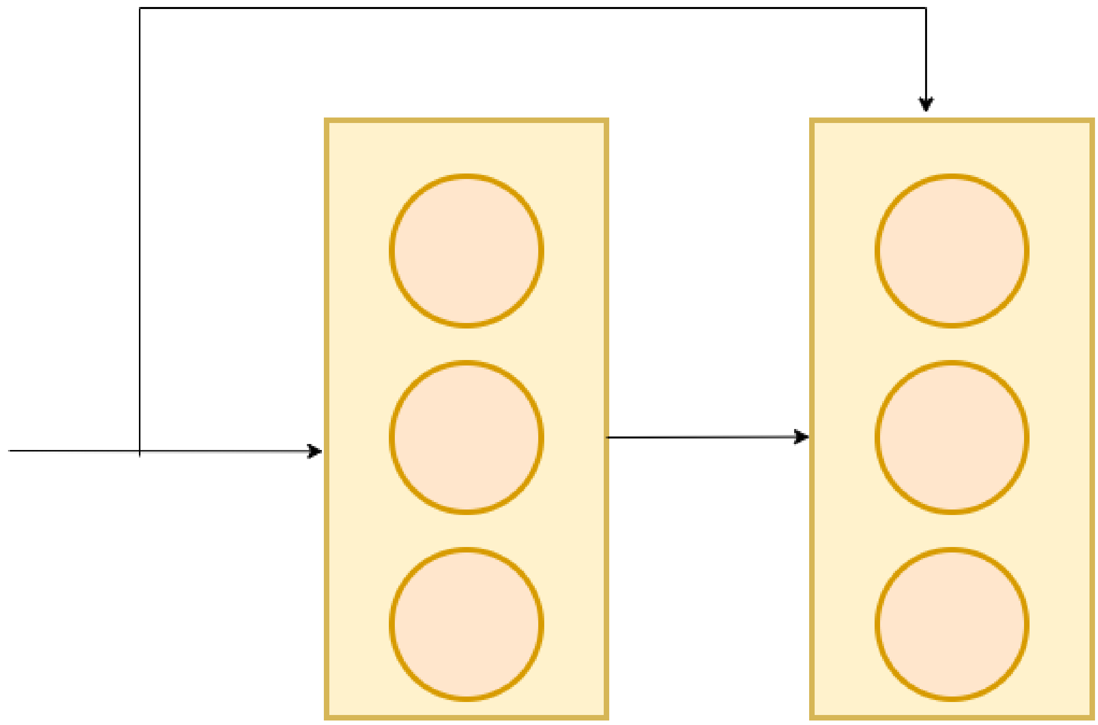

Figure 4.4: Skip Connections (Residual Networks)

In the traditional Networks, input only transfers one block to another in a series. However, in Residual Networks' Skip connection the input of the original network connects to the output of the convolutional block. Here some layers of low value layers are being skipped. Therefore, it solved the problem of vanishing gradient. The main idea of the Residual Network is to increase the overall accuracy. If we consider X is input and Y is output then, Y=F(X), where in ResNet the algorithm gets trained on the basis of F(X).

Finally, it should be noted that Inception-ResNet-v2 was trained using over a million photographs from the ImageNet database. The 164-layer network can categorize photographs into 1000 different object categories, such as keyboards, mice, pens, and other animals. As a result, the network has amassed a library of rich feature representations for a wide range of images. The photo input size for the network is 299 by 299 pixels.

Chapter 5

Model Implementation and Demonstration

5.1 Implementation

In our study, we used a customized dataset that was collected from various types of open sources for nine diseases. We implemented MobilenetV2 for making our task mobile friendly and tensor flow to carve up our images and transfigure them into an array. After implementation, we got 96.77% accuracy.

5.1.1 Input Data-Preprocessing

The custom dataset used in our research - described in the previous chapter - consisted of nearly two thousand readings distributed across nine skin diseases, including Melanoma, Acne, Basal cell carcinoma, Actinic keratosis, Viral Skin Infections, Bacterial Skin Infections, Deep dermal disorder, Tinea, and Eczema.

Once downloaded, the dataset comes in text format and can easily be opened using any text file reading software like Microsoft Excel. The data would look like the data shown in figure 5.1 The two columns in the dataset represent id(image id), Disease(disease name) respectively. As a result, the dataset connects the image id with the respective disease name in the same row. These two attributes are necessary for our proposed model for detecting and classifying skin diseases.

168	TSR_2_202	Basal cell carcinoma
169	TSR_2_203	Basal cell carcinoma
170	TSR_2_339	Bacterial Skin Infection
171	TSR_2_340	Bacterial Skin Infection
172	TSR_2_341	Bacterial Skin Infection
173	TSR_2_342	Bacterial Skin Infection
174	TSR_2_343	Bacterial Skin Infection
175	TSR_2_344	Bacterial Skin Infection
176	TSR_2_345	Bacterial Skin Infection
177	TSR_2_621	Viral skin infections
178	TSR_2_622	Viral skin infections

Figure 5.1: A sample of raw data as obtained from the custom dataset

5.1.2 Algorithm Implementation

Tensor Flow

TensorFlow offers several abstraction levels from which we may choose the one that best fits our requirements. To get started with TensorFlow and machine learning, use the high-level Keras API to construct and build models. In terms of flexibility, the eager implementation allows for quick iteration and straightforward debugging. Use the ADPI for decentralized training on diverse hardware configurations without modifying the model design for large ML training initiatives. Using Tensorflow architecture we have divided and split the pictures and converted them into an array.

For our research, we used google based classification architectures like Mobilenetv2, InceptionV3, RestNetV2 which work on CNN based algorithm called Inverted Residual Block, sometimes also called an MBConv Block. In preprocessing the dataset we established a bar plot from our input.csv file in figure 5.2 where all the Nine diseases and fresh skin and their total quantity is shown.

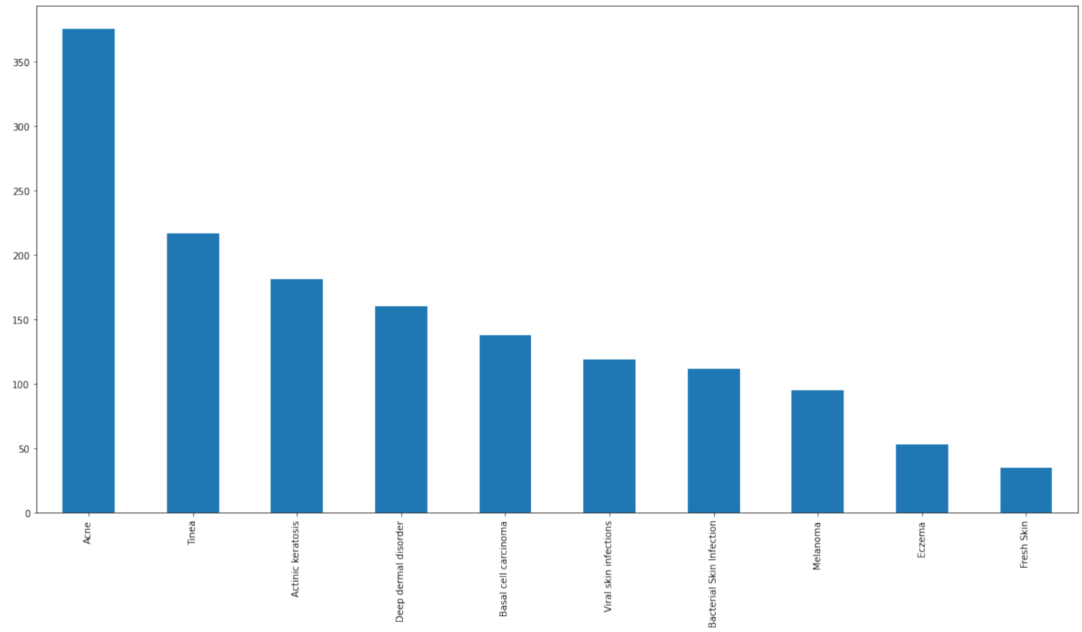

Figure 5.2: Input Data Bar Plotting in Pre-Processing

After preprocessing, we split our dataset into 80% train and 20% validation and then train our MobileNetV2, InceptionV3, and RestNetV2 model accordingly.

MobileNetV2

MobileNetV2 is a CNN design that tries to be mobile-friendly. It is a Convolutional Neural Network in complete form. It's a simple model that assists with image classification. It's based on an inverted residual structure, with bottleneck levels connected by residual connections. Lightweight depthwise convolutions are used as a source of non-linearity in the intermediate expansion layer filters.. Our MobileNetV2 framework used Convolutional Neural Network to process the split image and compare it to the referred image of disease from our dataset. However, in the MobileNetV2 framework, we have used the Keras library that was imported from the tensorflow hub and the model link:
https://tfhub.dev/google/imagenet/mobilenet_v2_130_224/classification/4

We had to send the model URL to our Create Model Function, which takes the URL as an input parameter, and then setup the model layers using the keras sequential technique, where we used our model URL in kerasLayer and set the activation value to "Softmax" in Dense Layer. The model is then compiled using Adam optimizer with loss function. Finally, we obtained the model summary when we finished building the model.

Moreover, we set patience value 3 of our TensorBoard callback function. And set the number of Epochs as 100.Lastly, we called our model train function where we fit the model with 80% training and 20% validation with the callbacks of early_stopping.And for the first time training we get the model accuracy 0.8963 and validation accuracy of 0.7150 for epoch 8 which are displayed below in Figure 5.3 :

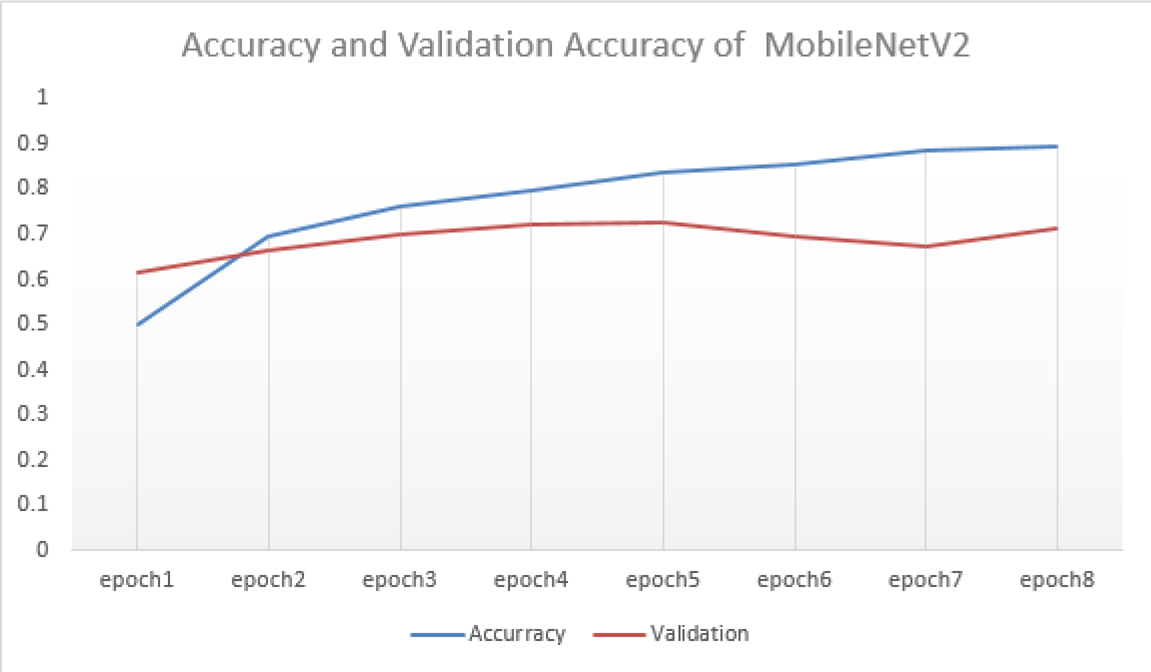

Figure 5.3: Accuracy and Validation Acurracy of MobileNetV2

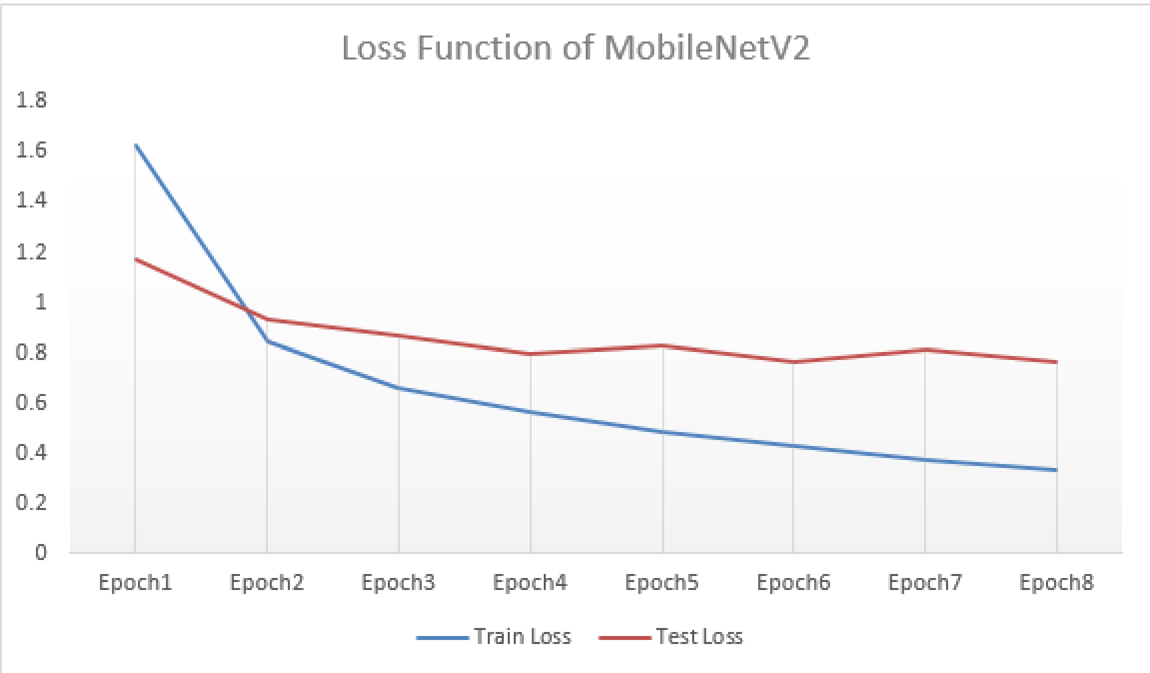

Figure 5.4: Loss function of MobileNetV2 Model

Lastly, we trained our model for a full data set to get higher accuracy and then we predicted our model for custom images to get real life accuracy results.

MobileNetV2 is widely used as the layers are lightweight and run time is less.Arguments that have been used in the MobileNetV2 are:-

1. **input_shape:** It is optional to provide a model with an input image resolution less than (224, 224, 3) when utilizing this shape tuple. A maximum of three input channels should be used (224, 224, 3). This parameter was used to resize the image, with the maximum size set at 224.

2. **alpha:** float type. In the MobileNetV2 specification, this is renamed as the width multiplier, but the term is kept for application compatibility. There is a MobileNetV1 model in Keras.

   (a) When alpha < 1, the number of filters in each layer decreases proportionally.

   (b) The number of filters in each layer rises proportionally as alpha ¿ 1.0.

   (c) If alpha $= 1.0$, each layer consists of the default amount of filters from the paper.

   However, we have used 3 layers in the model. Finally, we checked the validity of our algorithm by creating batches and then using the code : val_data= create_data_batches(X_val, y_val, valid_data=True)

3. **include_top:** Boolean type, it decides if the fully connected layer at the top of the network to include. Defaults to True.

4. **input_tensor:** Optional Keras tensor, for the model it has been used as the input image.

5. **pooling:** Its type is a string, when include_top is False then optional pooling mode for feature extraction happens.

6. **classes:** classes are Integer types, There is no limit to the number of classes that may be used to categorize pictures. It is provided only if the included top is True and if the weights argument is omitted.

7. **classifier_activation:** This is either a callable or str. The ”top” layer should get the activation function. This is ignored unless includes top=True. The logits of the ”top” layer are returned when classifier activation=None is specified. When loading pre-trained weights, the only valid values for classifier activation are None or ”softmax”[5].

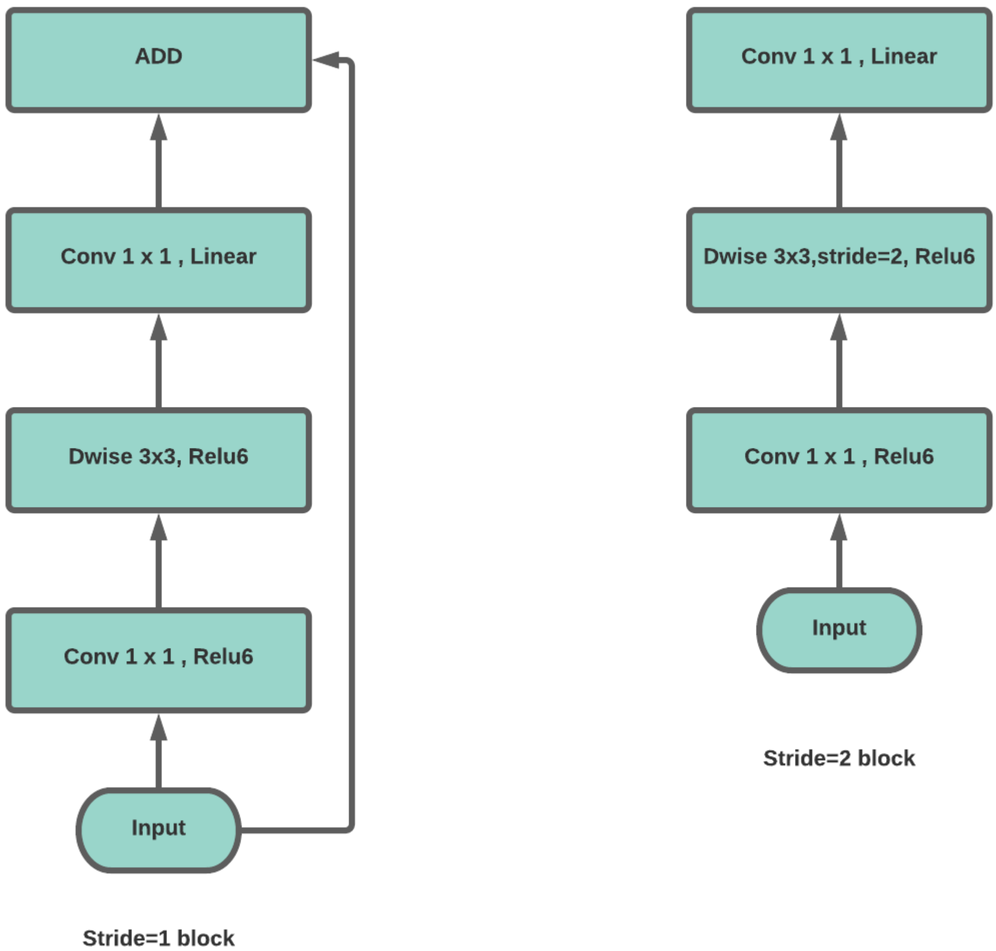

Figure 5.5: Flow diagram of MobileNetV2 Framework

The given whole arguments are based on the MobileNet V2 framework. However, we have called the MobileNet V2 framework and TensorFlow modules or libraries to handle the algorithm. Therefore, the combined usage of TensorFlow and MobileNetV2 helped us to achieve our expected output.

InceptionV3 and ResNetV2

Implementation of InceptionV3 and ResNetV2 in TensorFlow

In order to use TensorFlow Hub, the version of TensorFlow had to be greater or equal to 1.7, and we needed to install an additional package for TensorFlow Hub. Pre-trained models are managed as modules in TensorFlow Hub. At first we collected the URL of ResNetV2 and InceptionV3 from the TensorFlow hub and created the method named “create_method()” to pass the url as a parameter. By using Keras’s sequential approach we utilized a Keras sequential technique, where we used our model URL in the Keras layer and set the activation value to ”Softmax” in the Dense Layer. The model is then compiled using Adam optimizer with a loss function. And finally after building the model we got the summary of the model.

Therefore, setting patience value 3 of our TensorBoard callback function and setting the number of Epochs as 100.For, epoch 9 we get a 0.8850 accuracy rate with a 0.6600 validation rate for InceptionV3. In ResNetV2 for epoch 11, we acquire a model accuracy of 0.8737 and a validation accuracy of 0.6850 for the first time training. Lastly, Like MobileNetV3 we trained our model for a full data set to get higher accuracy, and then we predicted our model for custom images to get real-life accuracy results.

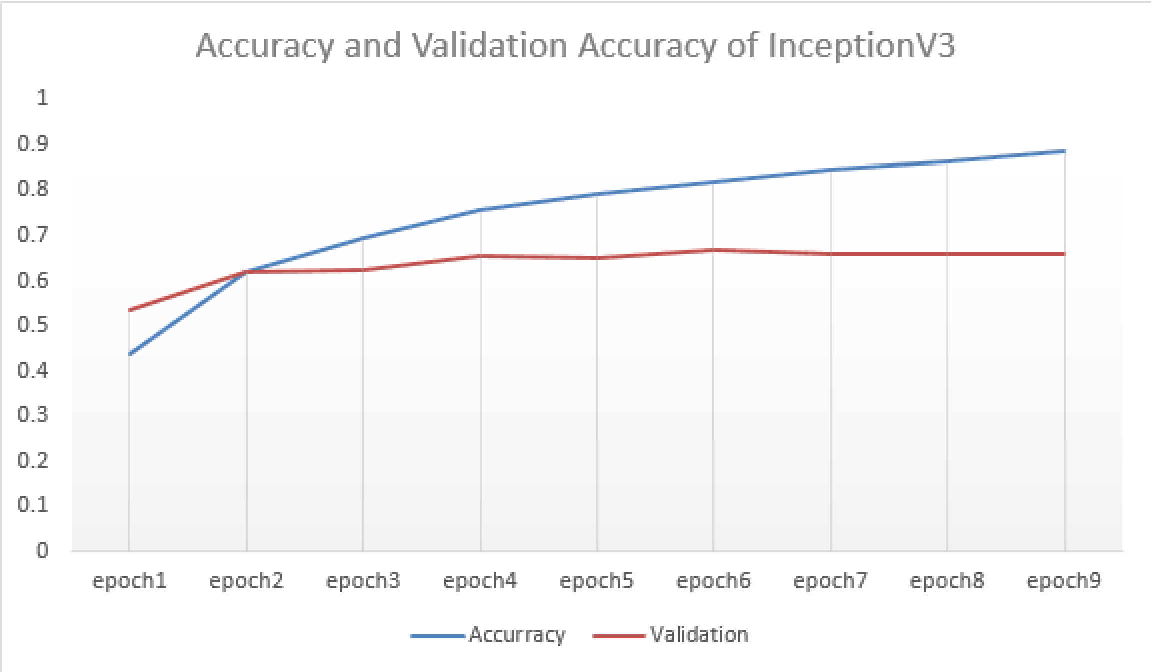

Figure 5.6: Accuracy and Validation Accuracy of InceptionV3 model

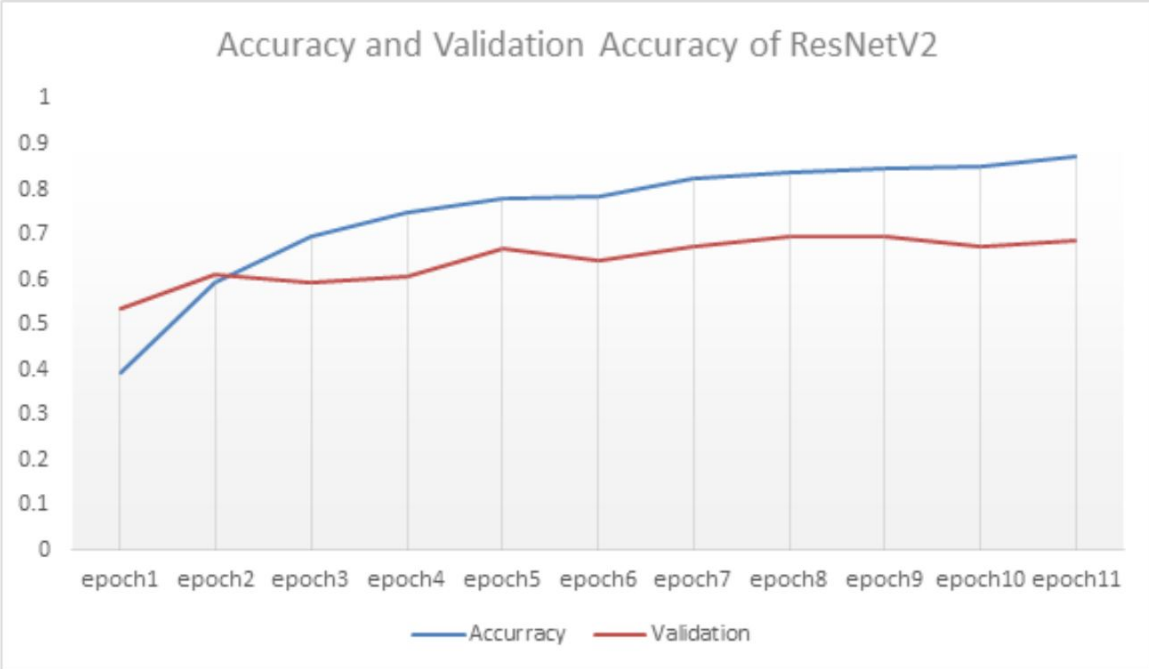

Figure 5.7: Accuracy and Validation in ResNetV2 model

Then, we finally fit the whole dataset for the model and predict for custom images.The arguments of the InceptionV3 and ResNetV2 are given below:

1. **input_tensor:** additional Keras tensor (output of layers.Input()) to utilize as image input for the model.

2. **weights:**None (random initialization), 'imagenet' (imagenet pre-training), or the location to the weights file to be loaded are all acceptable options. include_top: Here the layer is fully connected on the top of the model.

3. **classes:** undefined number of classes to categorize photos into, only if the included top is True and no weights parameter is supplied.

4. **pooling:** An optional pooling mode for feature extraction is employed when the included top is False.

   (a) None of this means that the model's output will be the last convolutional block's 4D tensor result.

   (b) avg indicates that the result of the last convolutional block will be subjected to global average pooling, resulting in a 2D tensor as the model's output.

   (c) indicates that global maximum pooling will be used. input_Shape: The input shape must be (224, 224, 3) (with 'channels last' data format) or (3, 224, 224) (with 'channels first' data format) if the included top is False. It should have three input channels with a minimum width and height of 32 pixels. For instance, (200, 200, 3) is a valid value.

5. **classifier_activation:** an str or a callable The activation function that should be used on the "top" layer. Unless included top=True, this is ignored. To

retrieve the logits of the "top" layer, set classifier activation=None. When importing pre-trained weights, the only options for classifier activation are None or "softmax."[9]

Chapter 6

Data Analysis, Comparison and Result

Here we are discussing the three distinct models that have been used to classify the images which are MobileNetV2, InceptionV3 and ResNetV2. Finally, after a detailed comparison, we come up with our optimum result.

6.1 Output of the models

For All the three models, we predict from our dataset with different values, and the highest similarity rate gets chosen. Figure 6.1 shows the similarity rate of all 9 diseases and fresh skin for MobileNetV2 and then the heighest percentage gets chosen. Meanwhile, after training the entire model, we reached an accuracy of 96.77% with a loss rate of 14.51% for MoblieNetV2, shown in figure 6.2

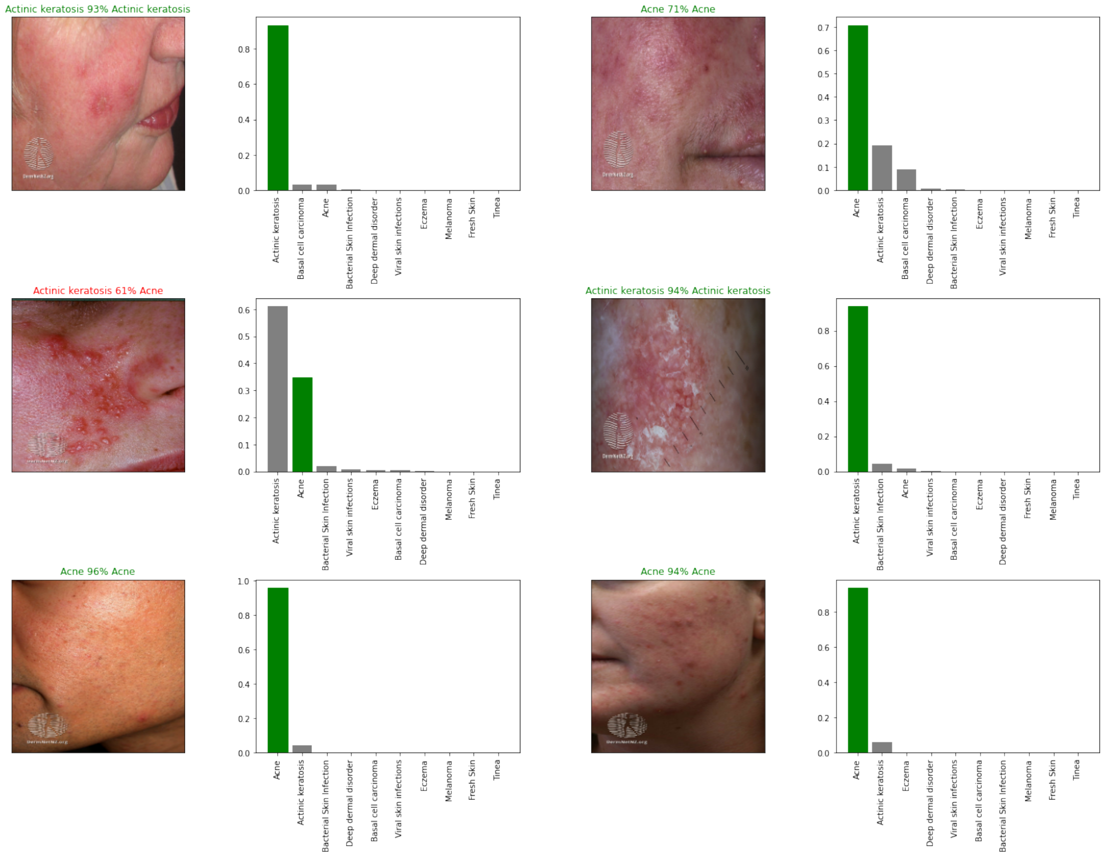

Figure 6.1: Similarity rate of all 9 diseases and fresh skin for MobileNetV2

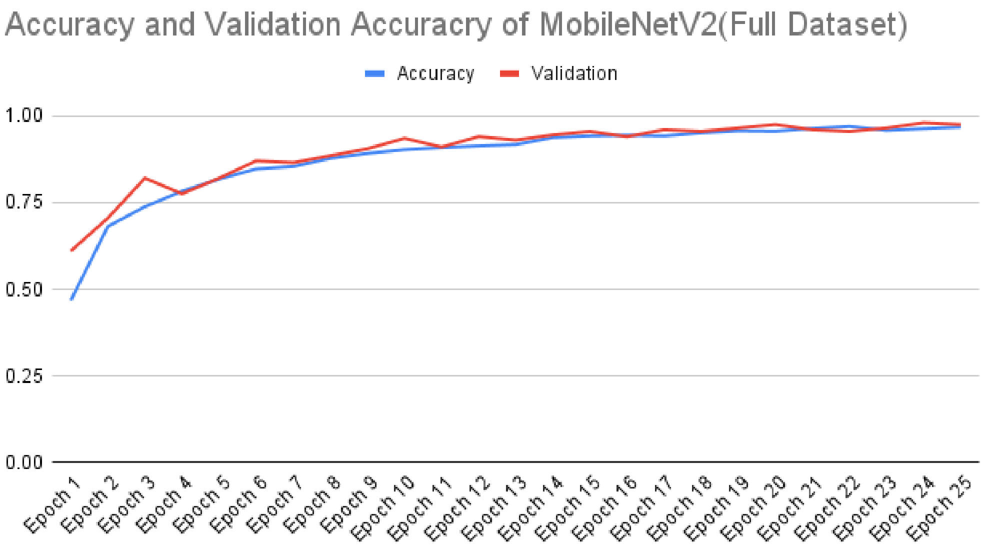

Figure 6.2: Accuracy rate and validation accuracy rate of MobileNetV2

Also, we get 0.1491 accuracy loss and 0.9677 validation accuracy loss in the 25th epoch. Loss function of MobileNetV2 for full dataset is shown in figure 6.3

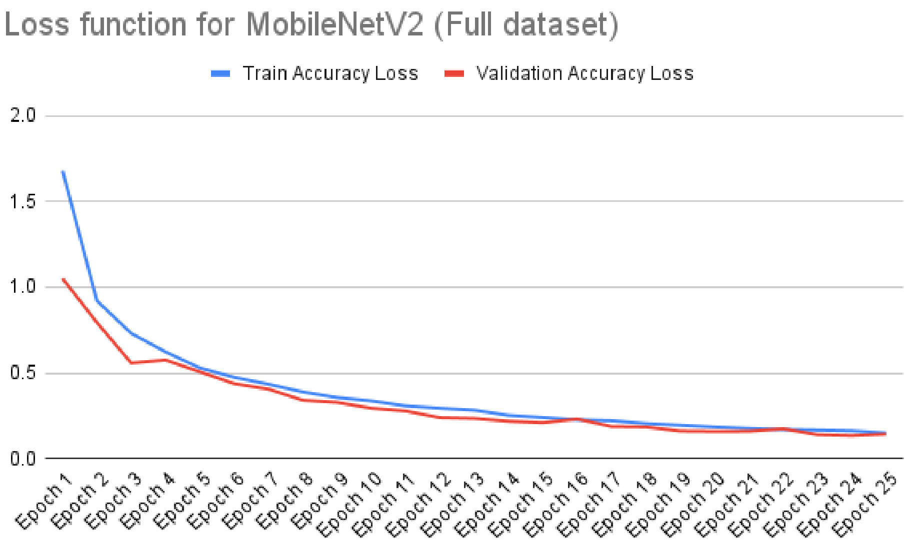

Figure 6.3: Loss function for MobileNetV2 (Full dataset)

Finally, to check practical usages, we collect random images of skin disease from the internet, which was not in the dataset before and store them in google drive. Then our trained MobileNetV2 model has successfully detected the disease from the custom images. Figure 6.4 shows us classifying diseases from custom images.

```
# Check custom image predictions
plt.figure(figsize=(20, 20))
for i, image in enumerate(custom_images):
  plt.subplot(1, 5, i+1)
  plt.xticks([])
  plt.yticks([])
  plt.title(custom_pred_labels[i])
  plt.imshow(image)
```

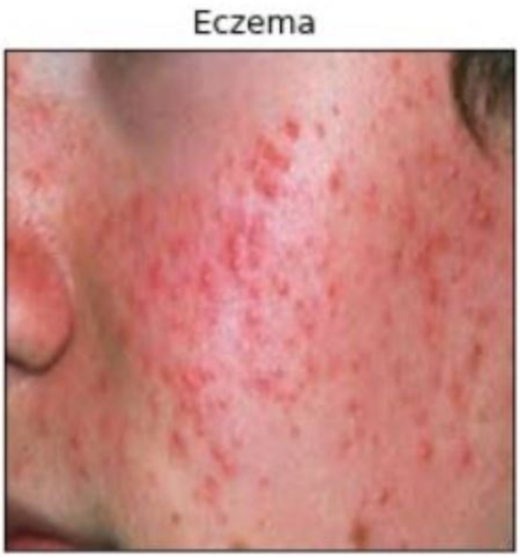

Figure 6.4: Classifying diseases from custom images

Meanwhile, for InceptionV3 and ResNetV2 after fitting the full dataset we get the accuracy 98.11% and 98.45% accordingly.

6.2 Detailed comparison

Convolutional neural networks such as InceptionV3, ResNetV2, and MobileNetV2 are often utilized for image categorization tasks. ResNet focuses on computational accuracy, whereas Inception emphasizes on computational cost. On the other hand MobileNet features the best version considering the weight as it is the lightweight model.

The Inception module performs many transformations on the same input map simultaneously before integrating the results into a single output. For each layer, it does a 5x5 convolution, 3x3 convolution, and max pooling, each of which transmits different information and is computationally costly. As a result, the designers of Inception decided to use dimension reductions to overcome the problem. Before going on to the bottlenecks of 3x3 and 5x5, dimension reduction refers to the usage of 1x1 convolution. It has a compressed form of spatial data as a consequence.

ResNet supported using network layers to fit a residual mapping rather than trying to match a desired underlying mapping directly. To put it another way, the network is attempting to learn F(x) + x rather than H(x). This eliminates the issue of fading gradients, which happens when the gradient signals of the error function drop exponentially as they are back propagated. The erroneous signals had become minimal by the time they reached the previous stratum. The gradient signal in ResNet might flow back to early layers via this "shortcut" mechanism, allowing for the creation of many layers of the network without sacrificing accuracy. The goal of MobileNet is to develop lightweight deep neural networks by using depthwise separable convolutions. The convolution kernel or filter is applied to all of the channels of the input image in a normal convolutional layer by doing a weighted sum of the input pixels with the filter, then sliding to the next input pixels across the pictures. Only the first layer of MobileNet employs standard convolution. The depth wise separable convolutions, which are a mix of depthwise and pointwise convolutions, are the following layers. The depthwise convolution performs the convolution independently for each channel. As a result, if the input picture contains three channels, the output image will have three channels as well. The input channels are filtered using depthwise convolution. The pointwise convolution is the following phase, which is identical to conventional convolution but with a 1x1 filter. The goal of pointwise convolution is to combine the depthwise convolution's output channels to produce additional features. As a result, the computing labor required is less than with traditional convolutional networks.

In our observation, for a full data set MobileNetV2, InceptionV3 and ResNetV2 have similar train accuracy and validation accuracy but among them figure 6.5 showed that MobileNetV2 has been the most consistent. Besides, as MobileNetV2 is a lightweight neural network, our experiment shows that MobileNetV2 has the lowest total params and also keras dense layers than InceptionV3 and ResNetV2. Figure 6.6 and 6.7 showed their comparison of total params and keras dense layers accordingly.

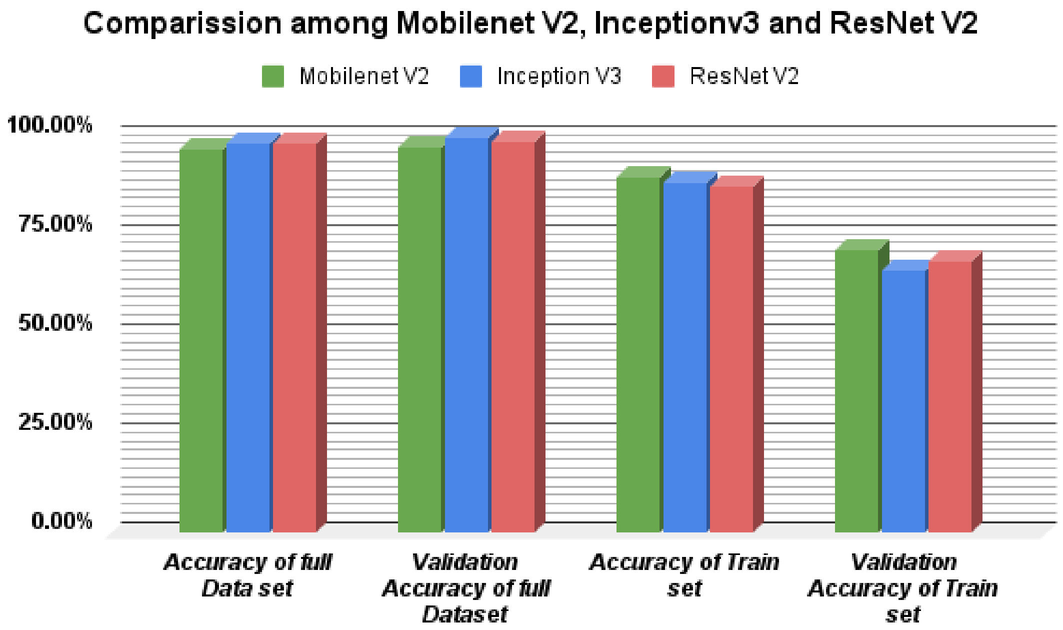

Figure 6.5: Comparison of accuracy rate among MobileNetV2, InceptionV3 and ResNetV2.

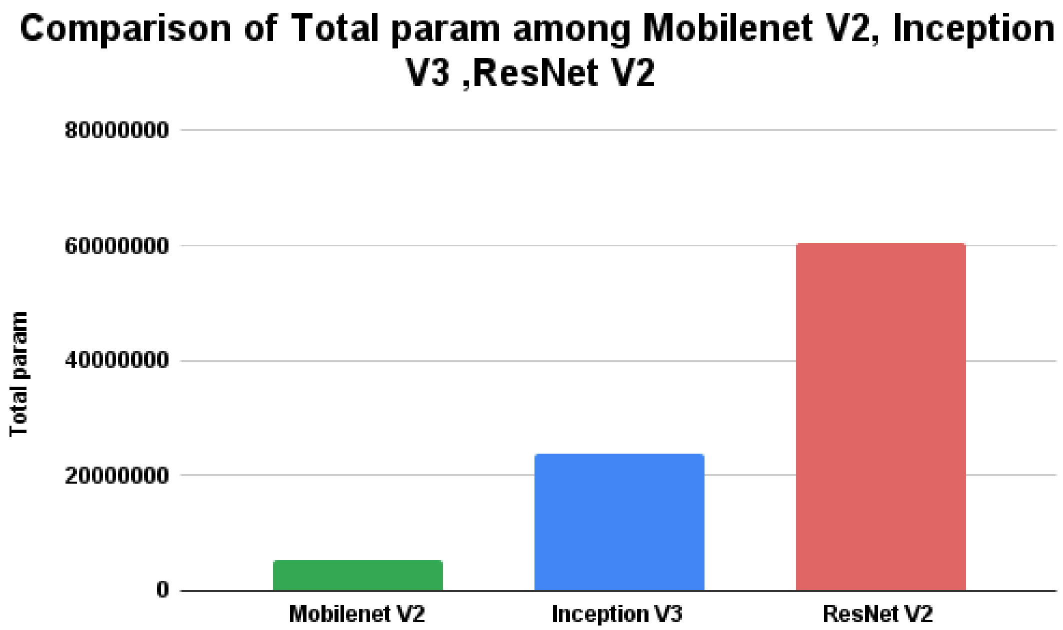

Figure 6.6: Comparison of Total param among MobileNetV2, InceptionV3 and ResNetV2.

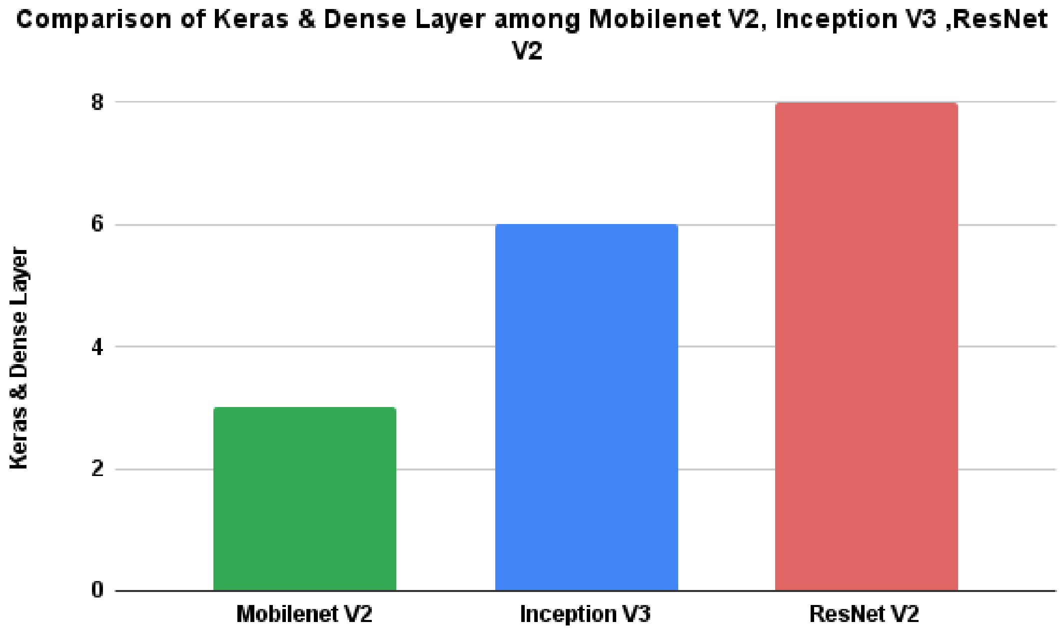

Figure 6.7: Comparison of Keras & Dense Layer among MobileNet V2, InceptionV3, and ResNet V2

Chapter 7

Conclusion

The goal of our future work is to create a mobile application that rural and urban people may use anywhere, at any time, to acquire a pre-medical evaluation and diagnose their skin ailment. This will not only save their lives but also make them more mindful of their skin since the initiative will prevent the spread of life-threatening skin illnesses. Thus we came up with a comparison between different models to determine the nine diseases along with the identification of fresh skin accurately. We utilized MobileNetV2, InceptionV3 and ResNetV2. After training the MobileNetV2 we are losing 33.56% accuracy. The training accuracy is 89.63% with 71.50% validation accuracy. After fitting the full model we are losing 14.91% accuracy and the final accuracy is 96.77% and 97.50% validation accuracy. On the other hand for InceptionV3 we are losing 38.92% accuracy. The training accuracy is 88.50% with 66.00% validation accuracy. After fitting the full model time we are losing 10.63% accuracy and the final accuracy is 98.11% with 99.50% validation accuracy. Again for ResNetV2, we are losing 38.91% accuracy. The training accuracy is 87.37% with 68.50% validation accuracy. After fitting the full model time we are losing 08.37% accuracy and the final accuracy is 98.45% with 98.50% validation accuracy. Finally, we are validating MobileNetV2 a Google-based categorization architecture that is based on the Inverted Residual Block, commonly known as an MBConv Block. The MobileNetV2 model, as seen in the graph, is the lightest and most mobile-friendly. Because it has fewer Keras and Dense layers. As a result, despite its close accuracy, with InceptionV3 and ResNetV2 the MobileNetV2 outperforms the competition for our specific requirement.The goal of our future work is to create a mobile application that rural and urban people may use anywhere, at any time, to acquire a pre-medical evaluation and diagnose their skin ailment. This will not only save their lives but also make them more mindful of their skin since the initiative will prevent the spread of life-threatening skin illnesses. Our long-term objectives include data augmentation, as our unique dataset is unbalanced, and the creation of an ensemble model, which will allow us to improve accuracy while also allowing the model to run optimally.

Bibliography

[1] M. Lin, Q. Chen, and S. Yan, "Network in network," *CoRR*, vol. abs/1312.4400, 2014.

[2] J. Brownlee, "Overfitting and underfitting with machine learning algorithms," Mar. 2016. [Online]. Available: https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/ (visited on 10/12/2021).

[3] B. Marques and D. Calheiros, *Dermatology atlas*, www.atlasdermatologico.com.br, May 2017. [Online]. Available: http://www.atlasdermatologico.com.br/ (visited on 02/25/2021).

[4] , *Dermis*, DermIS.net, Oct. 2018. [Online]. Available: https://www.dermis.net/dermisroot/en/home/index.html (visited on 02/19/2021).

[5] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Inverted residuals and linear bottlenecks: Mobile networks forclassification, detection and segmentation," Jan. 2018.

[6] P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, Aug. 2018. DOI: 10.1038/sdata.2018.161. [Online]. Available: https://www.nature.com/articles/sdata2018161 (visited on 05/25/2021).

[7] 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*, vol. 18, Jul. 2018. DOI: 10.1186/s12911-018-0631-9. (visited on 09/10/2021).

[8] S. Albawi, Y. Abbas, and Y. Almadanie, "Robust skin diseases detection and classification using deep neural networks," *International Journal of Engineering and Technology*, vol. 7, pp. 6473–6480, Jun. 2019. DOI: 10.14419/ijet.v7i4.24178.

[9] Fransiska, *The differences between inception, resnet, and mobilenet*, Medium, Aug. 2019. [Online]. Available: https://medium.com/@fransiska26/the-differences-between-inception-resnet-and-mobilenet-e97736a709b0 (visited on 10/11/2021).

[10] K. Sriwong, S. Bunrit, K. Kerdprasop, and N. Kerdprasop, "Dermatological classification using deep learning of skin image and patient background knowledge," *International Journal of Machine Learning and Computing*, vol. 9, pp. 862–867, Dec. 2019. DOI: 10.18178/ijmlc.2019.9.6.884.

[11] H. Janoria, J. Minj, and P. Patre, "Classification of skin disease from skin images using transfer learning technique," Nov. 2020, pp. 888–895. DOI: 10.1109/ICECA49313.2020.9297567.

[12] L.-F. Li, X. Wang, W.-J. Hu, N. N. Xiong, Y.-X. Du, and B.-S. Li, "Deep learning in skin disease image recognition: A review," *IEEE Access*, vol. 8, pp. 208 264–208 280, 2020. DOI: 10.1109/ACCESS.2020.3037258.

[13] T.-C. Pham, A. Doucet, C.-M. Luong, C.-T. Tran, and V.-D. Hoang, "Improving skin-disease classification based on customized loss function combined with balanced mini-batch logic and real-time image augmentation," *IEEE Access*, vol. 8, pp. 150 725–150 737, 2020. DOI: 10.1109/ACCESS.2020.3016653.

[14] N. Agrawal and S. Aurelia, "Corroboration of skin diseases: Melanoma, vitiligo amp; vascular tumor using transfer learning," in *2021 7th International Conference on Electrical Energy Systems (ICEES)*, 2021, pp. 590–592. DOI: 10.1109/ICEES51510.2021.9383682.

[15] S. Jiang, H. Li, and Z. Jin, "A visually interpretable deep learning framework for histopathological image-based skin cancer diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1483–1494, 2021. DOI: 10.1109/JBHI.2021.3052044.

[16] J. Puritt and M. E.Belec, "7 types of data bias in machine learning," Feb. 2021. [Online]. Available: https://www.telusinternational.com/articles/7-types-of-data-bias-in-machine-learning?INTCMP=ti_lbai (visited on 11/09/2021).