Diagnosis of Measles Disease Using Image Processing in children under

5 Afghan Children

Submitted by

Hamidullah Jaffari

Nowroz Bunyadi

Under the supervision of

Asst.Prof Ziaullah Momand

Diagnosis of Measles Disease Using Image Processing in children under 5 Afghan Children

Hamidullah Jaffari

Nowroz Bunyadi

A Thesis Submitted in Partial Fulfillment
of the Requirements for
the Degree of Bachelor in (Computer Science)
Faculty of Computer Science
Kabul University
2021

Thesis Committee

…………………………........................ Chairman of Thesis Committee

(name here.)

…………………………........................ Member and Thesis Advisor

(name here …)

…………………………........................ Member

(name here….)

…………………………........................ Member

(name here…..)

Abstract

ACKNOWLEDGEMENTS

CONTENTS

CONTENTS (Cont'd)

LIST OF TABLES

TABLE   **PAGE**

LIST OF FIGURES

FIGURE   **PAGE**

LIST OF SYMBOLS

SYMBOL		UNIT
A_{bi}	surface of the heat exchanger without considering the surface area of fin	m
A_{bo}	base surface of heat exchanger	m2
A_f	fin surface area	m2
A_p	corrected fin profile area	m2
C_c	heat capacity rate of cold fluid	W.K-1
C_h	heat capacity rate of hot fluid	W.K-1
C_{pc1}	constant pressure specific heat of dry air	J.kg-1.K-1
C_{pc2}	constant pressure specific heat of saturated water vapor	J.kg-1.K-1
hca	heat transfer coefficient of cold heat exchange	W.m-2.K-1
hha	heat transfer coefficient of hot heat exchange	W.m-2.K-1
I	electric current	amps
LA	length to area ratio of the TE element	m-1
L_{bc}	thickness of base cold heat exchanger	m
L_{bh}	thickness of base hot heat exchanger	m
L_c	corrected fin length	m
M	number of fins	
n	number density of the TE element	elements.m-2
P	electric power applied to TE modules	W
p	pressure	N.m-2
q_c	heat transfer rate at cold side	W
q_h	heat transfer rate at hot side	W

LIST OF TECHNICAL VOCABULARY AND ABBREVIATIONS

ANN	=	Artificial Neural Network
CNN	=	Convolutional Neural Network
DL	=	Deep Learning
DNN	=	Deep Neural Networks
DT	=	Decision Trees
KMC	=	K-Means Clustering
KNN	=	K-Nearest Neighbors
LR	=	Logistic Regression
ML	=	Machine Learning
NN	=	Neural Network
RF	=	Random Forest
SVM	=	Support Vector Machine

Chapter 1

1.1 Introduction

Measles is an acute viral respiratory illness caused by a single-stranded, enveloped RNA virus with 1 serotype, it is classified as a member of the genus Morbillivirus in the Paramyxoviridae family. Humans are the only natural hosts of the measles virus. Measles in history was considered to be a life event that almost all children went through. References to measles can be found as far back as the 7th century A.D. Measles was described by Muhammad ibn Zakariya ar-Razi (860-930) or Rhazes – a Persian philosopher and physician, in the 10th century A.D. as a disease that is "more dreaded than smallpox". Reza published a book entitled "The Book of Smallpox and Measles" [1]. Francis Home, a Scottish physician, demonstrated in 1575 that measles is caused by an infectious agent in blood patients.

This is endemic in Afghanistan, with almost all provinces reporting suspected cases every year. From 1 January 2021 to 29 January 2022, 35319 suspected measles cases were reported in Afghanistan. As Afghanistan currently ranks 6th in the world for the number of reported measles cases. Afghanistan does not have enough resources to establish a proper healthcare system that can keep up with the prevention and surveillance of measles.
As of now, there is a lack of essentials such as personal protective equipment (PPEs) with only enough to accommodate 5% of total confirmed cases. Alongside, there is a shortage of healthcare professionals (HCPs). That is why we have investigated and pursued many effective computerized ways to diagnose it. We have pursued and studied SVM and KNN, which is a machine learning algorithm, and CNN, which is a deep learning algorithm.

1.2 Research Goals

1. To make a robust and clean dataset from the real data in Afghanistan and evaluate the situation properly and accurately.
2. To develop predictive models using machine learning and deep learning to diagnose measles disease.
3. To evaluate the model and deploy it to real production.

1.3 Research Questions

The question we are going to cover in this research:

1. Which approach is the most efficient and accurate approach to diagnosing measles disease?
2. Which machine learning approach other people use to diagnose measles disease?

1.4 Problem Statement

Measles is a highly contagious infectious disease caused by the measles virus. This is endemic in Afghanistan, with almost all provinces reporting suspected cases every year. From 1 January 2021 to 29 January 2022, 35319 suspected measles cases were reported in Afghanistan [1]. There is not any system that clears how many people are infected with the disease or which region has the highest amount of infection. It is critical that we must be completely aware of the outbreak of the disease to prevent this from spreading. To accomplish that we must provide a system, so that people can easily register themselves to the system if they are infected.

The most common symptoms of measles include high fever, tiredness, cough, runny nose, and a red rash, which increase the similarity of measles to other contagious diseases and make it hard to diagnose it. As well as there is not any online system in the world to diagnose it. Because of the high similarity of measles to other diseases people spend every kind of drug and tablet to cure it. Unfortunately, instead of curing that disease, they cause other diseases. We have to search about this and find natural drugs which are affordable and accessible to them to cure it. We must guide them efficiently about what they must do if they are infected with the disease.

As we know that measles is a highly contagious disease, for example, the discovery shows that from ten people in one home one of them is infected measles will infect nine of them [2]. As we know it is hard to diagnose it through laboratory confirmation due to a very high degree of lack of resources in Afghanistan.

Chapter 2

Related work/Background

2.1 Introduction to Measles

2.1.1 Measles

Measles is a highly contagious, serious disease caused by a virus. Before the introduction of the measles vaccine in 1963 and widespread vaccination, major epidemics occurred approximately every 2–3 years and measles caused an estimated 2.6 million deaths each year.
More than 140 000 people died from measles in 2018 – mostly children under the age of 5 years, despite the availability of a safe and effective vaccine.

Measles is caused by a virus in the paramyxovirus family and it is normally passed through direct contact and through the air. The virus infects the respiratory tract, then spreads throughout the body. Measles is a human disease and is not known to occur in animals [4].

2.1.2 Common Symptoms

Symptoms of measles generally first appear within 10 to 12 days of exposure to the virus. They include:

1. cough
2. fever
3. runny nose
4. red eyes
5. sore throat
6. white spots inside the mouth

A widespread skin rash is a classic sign of measles. This rash can last up to 7 days and generally appears within 14 days of exposure to the virus. It commonly develops on the head and slowly spreads to other parts of the body [5].

2.1.3 People at high risk of complication

Common complications from measles include otitis media, bronchopneumonia, laryngotracheobronchitis, and diarrhea. People at high risk for complications are infants or children under the age of five, adults above the age of 20, pregnant women, and people with compromised immune systems [6].

2.1.4 Measles in Afghanistan

Measles is endemic in Afghanistan, with almost all provinces reporting suspected cases every year. Following periods of lower transmission in 2019 and 2020, and amid the humanitarian crisis in Afghanistan, weekly notifications of suspected measles cases have been increasing in all provinces since the end of July 2021, with the highest weekly toll observed in the last four weeks in January 2022 (Figure 1). The number of cases and deaths increased by 18% and 40% respectively from week 4, 2022 (week commencing January 24) to week 5, 2022 (week commencing January 31).

Figure 1. Weekly number of reported suspected measles cases in Afghanistan. Week 1, 2018 to week 4, 2022. Source: WHO Country Office in Afghanistan, WHO Regional Office for the Eastern Mediterranean.
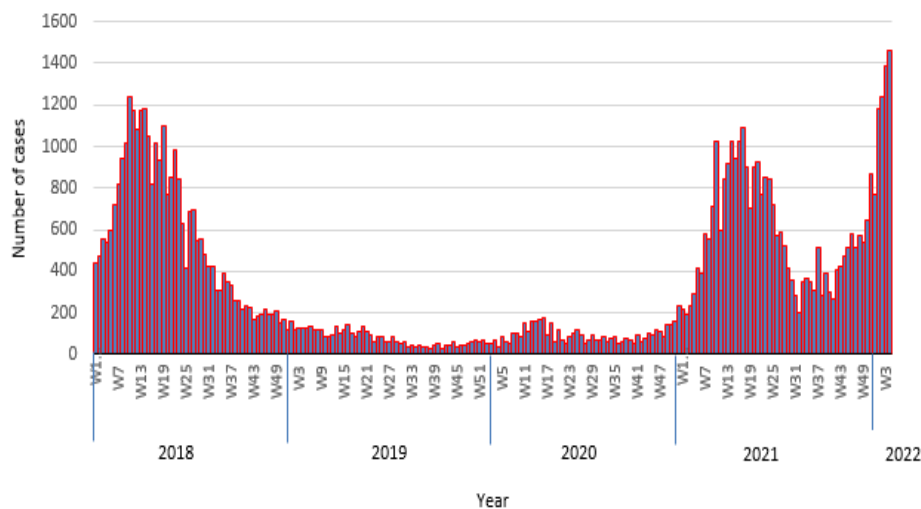

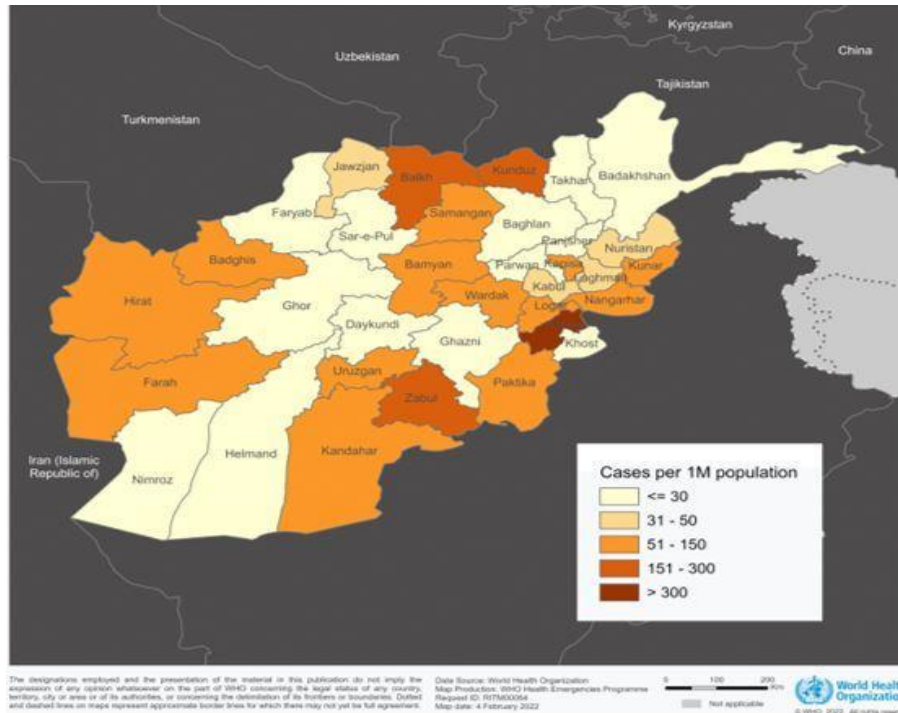Figure1.1Weekly number of reported suspected cases

Figure1.2 Number of reported cases by province

2.2 Related Work

Few researchers have proposed image processing-based techniques to detect measles diseases. Here we briefly review some of the techniques as reported in the literature.

As there are no public resources available that contain an extensive library of measles images specifically, they collected the data for their study using the Bing Web Search API (part of the Microsoft Azure package) to parse images from the web. The dataset that they have collected contains rash images of 11 different diseases containing Bowens, chickenpox, chigger, bites, dermatofibroma, eczema, enterovirus, keratosis, measles, psoriasis, ringworm, and scabies. Additionally, images of normal skin are also included in the dataset. They used transfer learning to develop their CNN model to detect measles rash through skin images. They also tried oversampling and image augmentation techniques using the Keras library because the two image classes in their dataset i.e., measles vs. non-measles, are highly imbalanced.

In the initial phase of model training, they kept all the convolutional layers, i.e. the backbone of ResNet-50, with their pre-trained weights and trained only the last few layers of the model. A stratified 5-fold cross-validation was conducted to train and evaluate the model. After the 5 iterations, the average performances of the models were computed that the average accuracy, sensitivity, and specificity of the model are 94.8%, 74.1%, and 97.6%, respectively [10].

The second study proposed a novel automatic deep learning-based classifier (CNN) to distinguish the distinctive HFMD rash from various other skin conditions. Their research showed that deep learning solutions could perform on par and even better methods on the tasks than classifiers to pick the color, texture manually, and shape features, even providing a minimal number of labeled images. Moreover, an advantage of their proposed method is that it does not need manual pre-processing of images. They have tested three pre-trained CNN models, ResNet (34, 50) and inception v3 on 2079 clinical images. In the experiment stage, we analyzed the performance of ResNet 50, 34, and inception v3, in which ResNet50 shows better performances than others. Their proposed model achieved the classification average accuracy of 0.954 and average sensitivity of 0.833 on the validation dataset [11].

In another research, which was about skin disease detection in Saudi Arabia, the method of detection was designed by using a a pre-trained convolutional neural network (AlexNet) and SWM. Their proposed system achieved an accuracy of 95% [12].

Another system proposed is a skin disease detection system. This system uses images of skin captured with a camera to detect if it is healthy or not; if not, then classified as Melanoma, Eczema, or leprosy.

The proposed system uses image processing and machine learning techniques. The process begins with pre-processing an input image using contrast enhancement and grayscale conversion. Global value thresholding techniques are used to segment the pre-processed images through which the actual affected region is obtained. They have used decision trees to classify the image.

Chapter 3

2.3 Methodology

2.3.1 Laboratory Diagnosis

Laboratory confirmation of measles is based on the detection of anti-MeV IgM antibodies or detection of MeV RNA by reverse transcription PCR (RT-PCR) in clinical samples. The most commonly used method for laboratory confirmation is the detection of IgM, usually by enzyme immunoassay, in serum samples collected at first contact with a suspected case [7].

RT-PCR, which is having an increasing role in case confirmation, has the highest sensitivity if samples are collected as early as possible after the onset of rash. In addition to the throat or nasal swabs, other clinical specimens that can be used for RT-PCR include oral fluid, urine, and peripheral blood mononuclear cells [8], [9].

Challenges of Laboratory Diagnosis

The member states of the WHO European Region adopted the goal of measles and rubella elimination more than 10 years ago, but so far only 21 of 53 countries have reached this target. Laboratory investigation of the suspected cases is essential to support disease elimination efforts. Therefore, WHO maintains a network of accredited laboratories providing high-quality testing, Laboratory investigation heavily relies on specific IgM serology and increasingly on virus detection by reverse transcription (RT)-PCR, but other methods such as lgG avidity testing and genetic characterization of virus strains have gained in importance. In elimination settings, often few samples from suspected cases are available for testing, but testing proficiency must be maintained. Thus, it is necessary to perform comprehensive and potentially time-consuming, and costly investigations of every suspected case using quality-controlled laboratory methods [13].

Challenges of Laboratory Diagnosis in Afghanistan

As Afghanistan currently ranks 6th in the world for the number of reported measles cases. Afghanistan has made great progress in its efforts to eradicate measles with reported vaccine coverage of 79% in 2010 from 42% in 2000, which has been attributed to a mass campaign of 2001 -2002.
Currently the country's healthcare service provider "Sehatmandi" is not functioning properly, which will result in the resurgence of many infectious diseases including measles. Afghanistan does not have enough resources to establish a proper healthcare system that can keep up with the prevention and surveillance of measles.
As of now there is a lack of essentials such as personal protective equipment (PPEs) with only enough to accommodate 5% of total confirmed cases. Alongside, there is a shortage of healthcare professionals (HCPs); a ratio of 9.4 HCPs and 1.9 physicians per 10 000 individuals. Moreover, those with measles will be unable to receive timely healthcare services, especially in areas with high conflict, and poor availability of healthcare services such as rural areas [14].

2.3.2 Computerized Diagnosis

Consider that we are living in a place that is far away from a hospital or do not have sufficient money to cover up the hospital bill or do not have enough time to take off work. In such cases, the disease diagnosis through sophisticated machines would be lifesaving. Scientists had developed numerous artificially intelligent diagnosis algorithms for detecting various diseases, Heart Diseases, Diabetic Diseases, liver Diseases, Skin diseases, Parkinson's Diseases, etc. Deep learning uses large artificial neural network layers having interconnected nodes that can rearrange themselves as and when new information comes in. This technique allows the computers to self-learn on their own without the need for human programming. Machine Learning (ML) is used practically everywhere, from cutting-edge technology (such as mobile phones, computers, and robotics) to health care (i.e., disease diagnosis, safety). ML is gaining traction in various fields, including disease diagnosis in health care. Many researchers and practitioners illustrate the promise of Machine Learning-based Disease Diagnosis (MLBDD), which is inexpensive and time-efficient.

3.2.1 Machine learning Methods

Machine Learning (ML) is an approach that analyzes data samples to create main conclusions using mathematical and statistical approaches, allowing machines to learn without programming. Arthur Samuel presented machine learning in games and pattern recognition algorithms to learn from experience in 1959, which was the first time the important advancement was recognized. The core principle of ML is to learn from data in order to forecast or make decisions depending on the assigned task. Thanks to Machine Learning (ML) technology, many time-consuming jobs may now be completed swiftly and with minimal effort. With the exponential expansion of computer power and data capacity, it is becoming simpler to train data-driven machine learning models to predict outcomes with near-perfect accuracy. Several papers offer various sorts of ML approaches [15]. ML algorithms can be divided into several subgroups based on different learning approaches, as shown in Fig. 1. Some of the popular ML algorithms include Linear Regression, Logistic Regression, Support Vector Machines (SVM), and Random Forest (RF), and Naïve Bayes (NB).

SVM (Support Vector Machine)

Support Vector Machine (SVM) was first heard in 1992, introduced by Boser, Guyon, and Vapnik in COLT-92. Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression [16].

Recently, particular attention has been dedicated to Support Vector Machines as a classification method. SVMs have often been found to provide better classification results than other widely used pattern recognition methods, such as the maximum likelihood and neural network classifiers (Melgani and Bruzzone 2004, Theodoridis and Koutroumbas 2003). Thus, SVMs are very attractive for the classification of remotely sensed data.

For classification and regression-related challenges, Support Vector Machine (SVM) is a popular ML approach. SVM was introduced by Vapnik in the late 20th century. Apart from disease diagnosis, SVMs have been extensively employed in various other disciplines, including facial expression recognition, protein fold, distant homology discovery, speech recognition, and text classification. For unlabeled data, supervised ML algorithms are unable to perform. Using a hyperplane to find the clustering among the data, SVM can categorize unlabeled data. However, SVM output is not non-linearly separable. To overcome such problems, selecting appropriate kernels and parameters are two key factors when applying SVM in data analysis [17].

SVM for Classification

SVM is a useful technique for data classification. Even though it's considered that Neural Networks are easier to use than this, however, sometimes unsatisfactory results are obtained. A classification task usually involves training and testing data which consists of some data instances. Each instance on the training set contains one target value and several attributes. The goal of SVM is to produce a model which predicts the target value of data instances in the testing sets which are given only the attributes. Classification in SVM is an example of Supervised Learning. Known labels help indicate whether the system is performing in the right way or not. This information points to the desired response, validating the accuracy of the system, or being used to help the system learn to act correctly. A step in SVM classification involves identification as which are intimately connected to the known classes. This is called feature selection or feature extraction. Feature selection and SVM classification together have use even when the prediction of unknown samples is not necessary. They can be used to identify key sets which are involved in whatever processes distinguish the classes [18].

Proposed Procedure

Many beginners use the following procedure Now:
- Transform data to the format of an SVM package
- Randomly try a few kernels and parameters
- Test

We propose that beginners try the following procedure First:
- Transform data to the format on an SVM package
- Conduct simple scaling on the data
- Consider the RBF kernel $K(x, y) = e^{-\gamma \|x - y\|}$
- Use cross-validation to find the best parameter C and γ

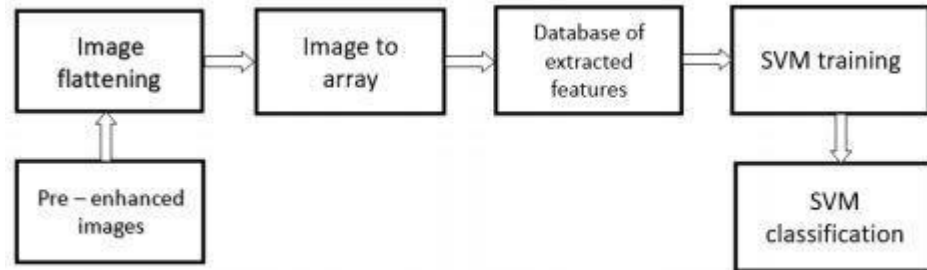- Use the best parameter C and γ to train the whole training set
- Test

Figure 2.3 Step for Working with SWM

Strength and Weakness of SVM:

The major strengths of SVM are that the training is relatively easy. No local optimal, unlike in neural networks. It scales relatively well to high dimensional data and the trade-off between classifier complexity and error can be controlled explicitly. The weakness includes the need for a good kernel function [20].

3.2.1.2 KNN

It is the simplest of the classification algorithms. The algorithm takes advantage of previously classified data. Attribute vectors must be created to implement the algorithm. The k parameter in this algorithm specifies the number of neighborhoods. According to the specified k parameter, each data is assigned to the nearest neighbor [21].
In the field of module identification, the K-Nearest-Neighbor algorithm (KNN) is a non-parametric statistical approach that is mainly used for classification and regression. As one of the simplest algorithms among all the machine learning algorithms, the KNN algorithm is an instance-based algorithm that is very simple and efficient, it is a non-learning algorithm [22].

Figure 2. 4: KNN Classifier

KNN has a lot of different names such as K-Nearest Neighbors, Memory-Based Reasoning, Example-Based Reasoning, Instance-Based Learning, Case-Based Reasoning, and Lazy Learning.
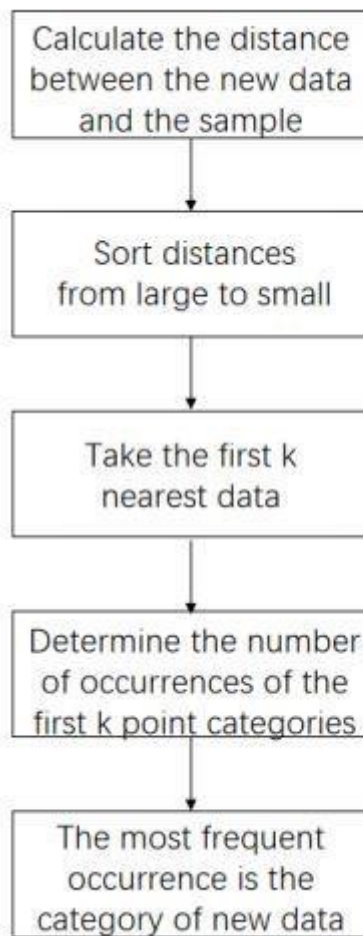
Figure 2.5 Operation flow of KNN algorithm

When should we use the KNN:

- We have label data.
- Data is error-free
- The dataset's size is small.

KNN Pros and Cons

Pros

- Easy to implement and understand because it does not include any assumptions. It is also heuristic.
- Responds quickly to changes in input during real-time usage.
- Can be easily applied to multi-class classification problems.

Cons

- As the data set grows, the speed of the algorithm decreases. Also, it becomes difficult to reach the output when the number of variables increases.
- Not capable of dealing with missing values and it is affected by outliers.
- To work properly, the variable features must be expressed in the same scale.
- Irrelevant or correlated features have high impact and must be eliminated.
- Typically, it is difficult to handle high dimensionality.
- Computational costs: memory and classification time computation.

2.3.2.2   Deep Learning Methods

Deep learning and Deep in Deep Learning

Deep learning is a specific subfield of machine learning: a new take on learning representations from data that puts an emphasis on learning successive layers of increasingly meaningful representations. The deep in deep learning isn't a reference to any kind of deeper understanding achieved by the approach; rather, it stands for this idea of successive layers of representations. How many layers contribute to a model of the data is called the depth of the model. Other appropriate names for the field could have been layered representations learning and hierarchical representations learning. Modern deep learning often involves tens or even hundreds of successive layers of representations— and they've all learned automatically from exposure to training data. Meanwhile, other approaches to machine learning tend to focus on learning only one or two layers of representations of the data; hence, they're sometimes called shallow learning. In deep learning, these layered representations are (almost always) learned via models called neural networks, structured in literal layers stacked on top of each other [23].

CNN

Convolutional neural networks are a specialized type of artificial neural network that uses a mathematical operation called convolution in place of general matrix multiplication in at least one of their layers [24]. In the past few decades, Deep learning has proved to be a very powerful tool because of its ability to handle large amounts of data. The interest to use hidden layers has surpassed traditional techniques, especially in pattern recognition. One of the most popular deep neural networks is the Convolutional Neural network.
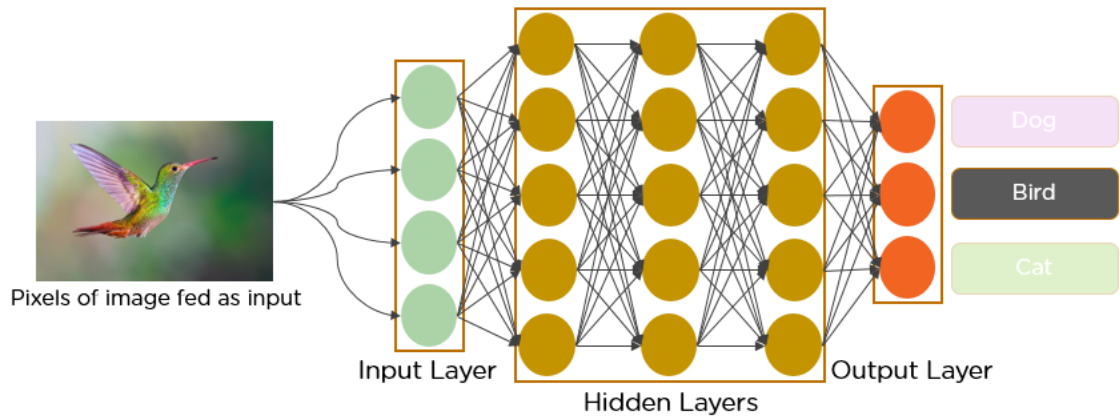
Figure 2.5 Neural Network Structure composed of layers and nodes which connected

In deep learning, a convolutional neural network is a class of deep neural networks, most commonly applied to analyze imagery. Now when we think of a neural network, we think about matrix multiplications but that is not the case with ConvNet. It uses a special technique called Convolution. Now in mathematics convolution is a mathematical operation on two functions that produces a third function that expresses how the shape of one is modified by the other.
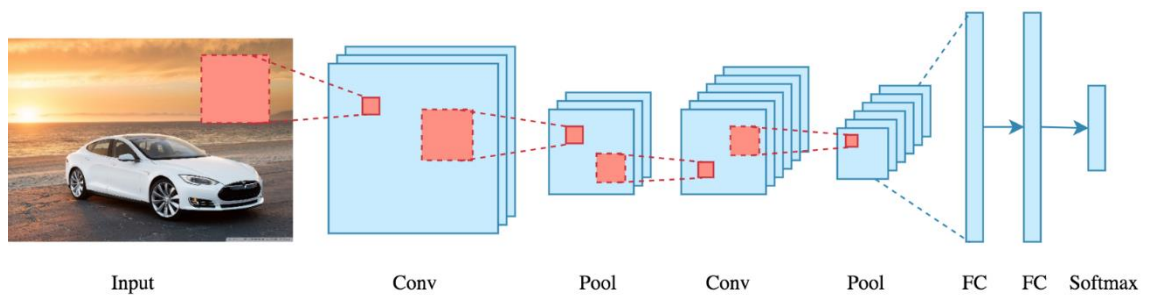

Figure 2.6 shows the modification of the shape

History of CNN

CNN was first developed and used around the 1980s. The most that a CNN could do at that time was recognize handwritten digits. It was mostly in the postal sectors to read zip codes, in codes, etc. The important thing to remember about any deep learning model is that it requires a large amount of data to train and also requires a lot of computing resources. This was a major drawback for CNNs at that period hence CNNs were only limited to the postal sectors and it failed to to enter the world of machine learning [25].
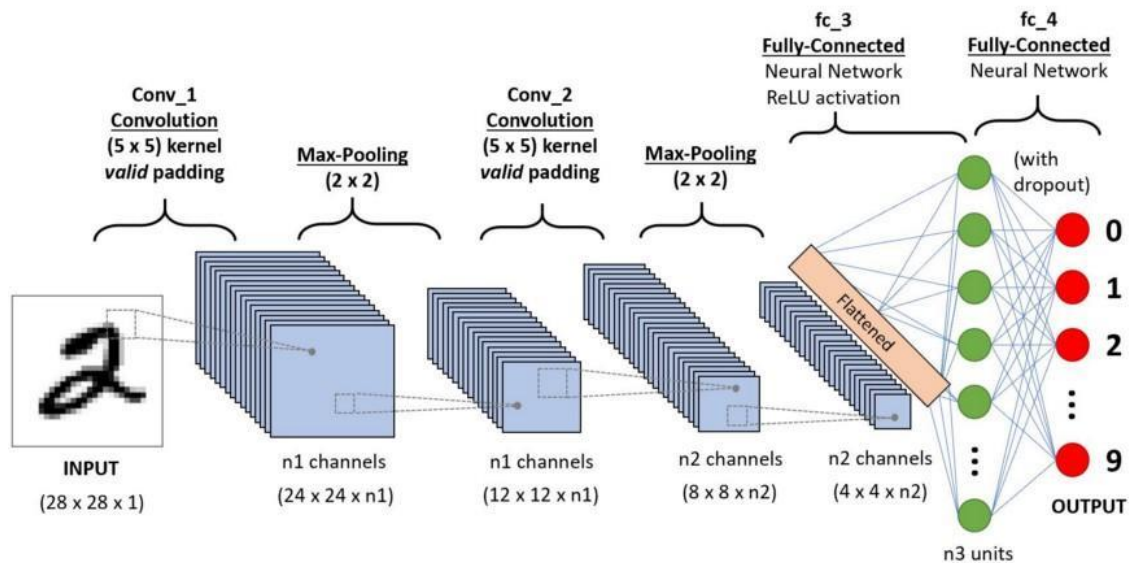
Figure 2.7 First period of CNN

CNN layer

A convolutional neural network consists of an input layer, hidden layers and an output layer. In any feed-forward neural network, any middle layers are called hidden because their inputs and outputs are masked by the activation function and final convolution. In a convolutional neural network, the hidden layers include layers that perform convolutions. Typically, this includes a layer that performs a dot product of the convolution kernel with the layer's input matrix. This product is usually the Frobenius inner product, and its activation function is commonly ReLU. As the convolution kernel slides along the input matrix for the layer, the convolution operation generates a feature map, which in turn contributes to the input of the next layer. This is followed by other layers such as pooling layers, fully connected layers, and normalization layers [26].

Convolutional Layer

In a CNN, the input is a tensor with a shape: (number of inputs) x (input height) x (input width) x (input channels). After passing through a convolutional layer, the image becomes abstracted to a feature map, also called an activation map, with shape: (number of inputs) x (feature map height) x (feature map width) x (feature map channels). Convolutional layers convolve the input and pass its result to the next layer. This is similar to the response of a neuron in the visual cortex to a specific stimulus. Each convolutional neural processes data only for its receptive field. Although fully connected feedforward neural networks can be used to learn features and classify data, this architecture is generally impractical for larger inputs such as high-resolution images. It would require a very high number of neurons, even in a shallow architecture, due to the large input size of images, where each pixel is a relevant input feature. For instance, a fully connected layer for a (small) image of size 100 x 100 has 10,000 weights for each neuron in the second layer. Instead, convolution reduces the number of free parameters, allowing the network to be deeper [27].

Pooling layer

The pooling layer will then simply perform down-sampling along the spatial dimensionality of the given input, further reducing the number of parameters within that activation [28]. Intuitively, the exact location of a feature is less important than its rough location relative to other features. This is the idea behind the use of pooling in convolutional neural networks. The pooling layer serves to progressively reduce the spatial size of the representation, reduce the number of parameters, memory footprint and amount of computation in the network, and hence to also control overfitting. This is known as down-sampling. It is common to periodically insert a pooling layer between successive convolutional layers (each one typically followed by an activation function, such as a ReLU layer) in a CNN architecture [27].

Fully-Connected Layer

The fully-connected layers will then perform the same duties found in standard ANNs and attempt to produce class scores from the activations, to be used for classification. It is also suggested that ReLu may be used between these layers, as to improve performance [28].

Image Preprocessing

Image data preprocessing is one of the most under-explored problems in the data science community. Every developer has a unique way of doing it. Some of the tools and platforms used in image preprocessing include Python, Pytorch, OpenCV, Keras, TensorFlow, and Pillow.
Image preprocessing is an important step and it is required because an image may contain many noises such as dermoscopic gel, air bubbles, and hairs.
Unfortunately, a few problems associated with image data include complexity, inaccuracy, and inadequacy. This is why before building a computer vision model, it is essential that the data is preprocessed (cleaned and processed to the desired format) to achieve that desired results. Finally, there are many ways of image preprocessing such as data augmentation, data enrichment, Grayscale conversion, normalization, and image standardization [29].
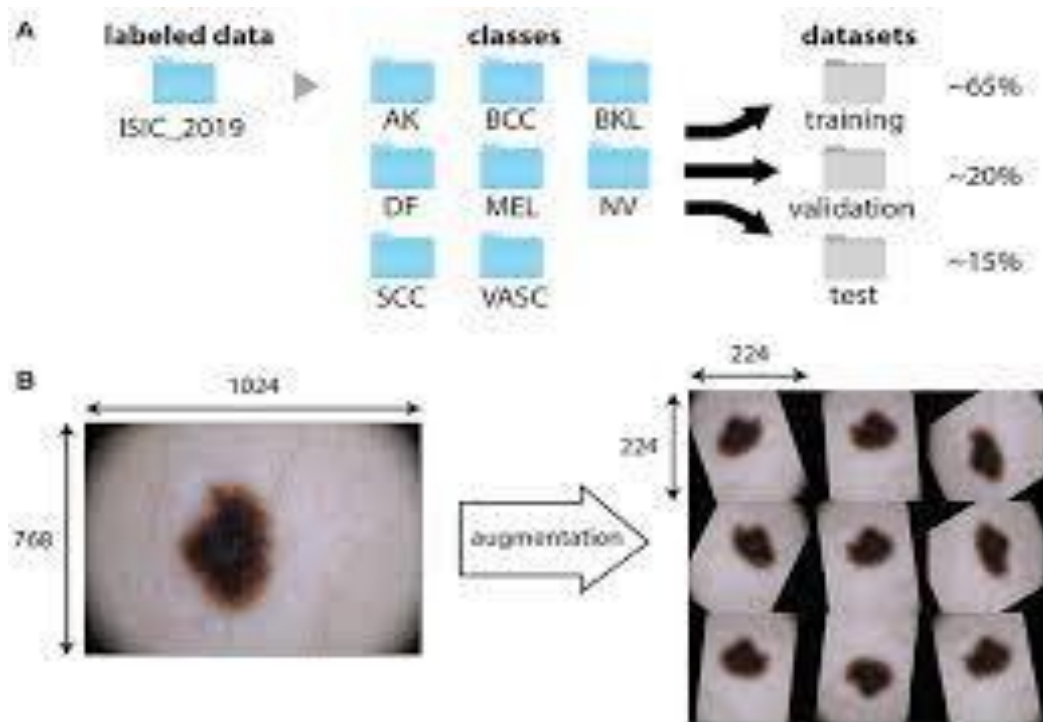
Figure 2.8 Augmentation

Image Segmentation

To build an effective deep learning-based classification system for detecting mild DED, we need to consider the importance of the architecture of the network as well as the importance of input data. To obtain an efficient result, input images play a a significant role. In retinal fundus images, variability such as the number of images, luminosity, contrast, and anatomical features determines the forthcoming result of the automatic disease detection algorithm. Therefore, feature segmentation enhances the value of the images for classification and contributes for better accuracy [30].
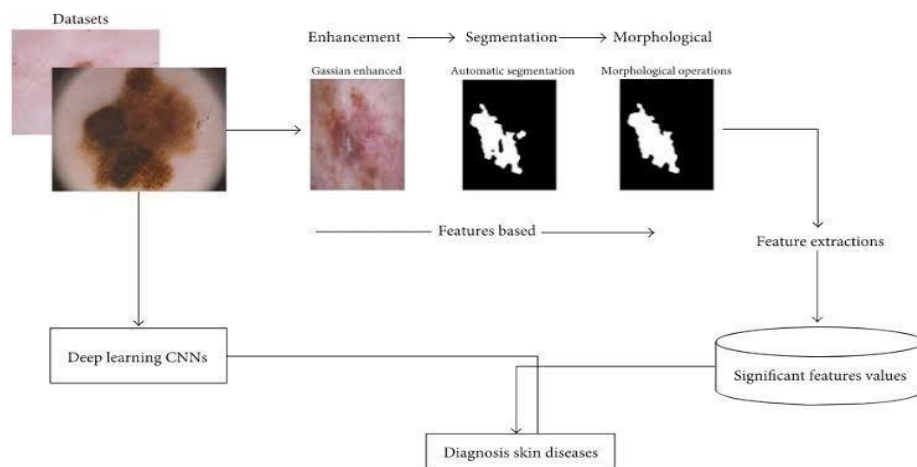

Figure 2.8 Segmentation

In computer vision, image segmentation is a way of segregating a digital image into multiple regions according to the different properties of pixels. Unlike classification and object detection, it is typically a low-level or pixel-level vision task as the spatial information of an image is very important for segmenting different regions semantically. Segmentation aims to extract meaningful information for easier analysis. In this case, the image pixels are labeled in such a way that every pixel in an image share certain characteristics such as color, intensity, texture, etc.

Mainly, image segmentation is of two types: semantic segmentation and instance segmentation. Also, there is another type called panoptic segmentation, which is the unified version of two basic segmentation processes [31].
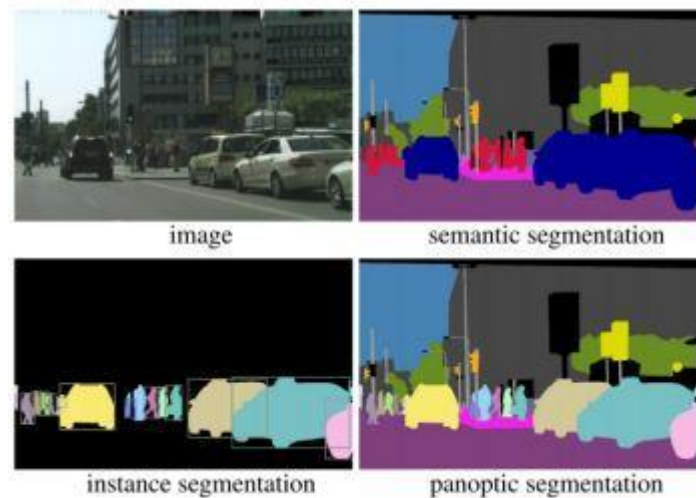

Figure 2.10 An example of different types of image segmentation

Feature extraction

In practice, two approaches are employed to reduce the dimensionality of datasets. These are, namely feature selection and feature extraction. Feature selection involves the selection (but not transformation) of features using certain optimization functions, whereas feature extraction allows transformation of features. More specifically, feature selection finds a subset of optimal features, whereas feature extraction creates a subset of new features by combinations of the existing features. Obviously, feature selection is one of the preprocessing techniques in data mining. Several methods exist in literature for feature selection. In practice, the methods are grouped into three categories, namely filter, wrapper, and hybrid model [32].

Filter Method

The filter method relies on general characteristics of data to evaluate and select feature subsets without involving any mining algorithm. These are typically faster and give an average accuracy for all the classifiers.

Wrapper Method

The Wrapper model requires one predetermined mining algorithm and uses its performance as the evaluation criteria. The wrapper methods can result in higher classification accuracy than filter methods for particular classifiers but they are less cost effective.

Embedded models

These models use the algorithms that adopt built-in feature selection methods. For instance, principal component analysis (PCA) and random forest have their own feature selection methods. These approaches tend to be between filters and wrappers in terms of computation complexity.

Feature selection is an essential part of machine learning since the performance of a machine learning model is much dependent on this phase.
The dataset was finally passed through this phase to select the features for the predictive model. For clinical purposes, it is arguable that parsimony is a desirable feature of a good predictive model.

Classification

Classification is a supervised learning approach for machine learning tasks. It requires a labeled dataset to map the data into specific groups or classes. There are various classification algorithms used to classify the skin disease images such as support vector machine, feed forward neural network, back propagation neural network, k-nearest neighbor, decision trees, etc.

Deep Learning is a part of machine learning algorithm inspired by the structure and function of the human brain commonly known as neural networks. Convolution Neural Networks (CNN) is a class of deep learning algorithms which is mostly used for analyzing the visual contents such as images and videos. With the development of CNN, there has been dramatic improvement observed to solve many classification-based problems in medical image analysis [33].
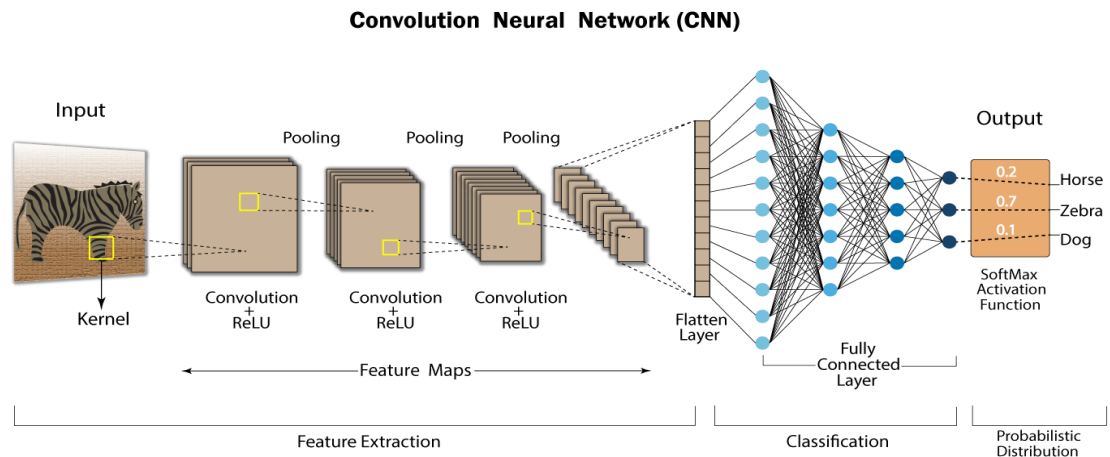
Convolution Neural Network (CNN)

Figure 2.11 shows the classification process

2.4 Description of Dataset

The dataset that we have collected contains three skin conditions such as Measles, Melanoma, and photos of people who are allergic to drugs. Our dataset contains about 300 hundred images.

Measles

Most of the measles images are collected from the Children Health hospital in Wazir Akbar Khan. But, some of them are similar photos to kid's measles, which is collected from some trusted and valid websites. We have gone two times in two weeks but, there were not a lot of patients with measles, so that is why we have not collected a lot of measles photos. We hope in the next semester we increase it.

Melanoma

Melanoma is one of the skin conditions, and more similar to measles, and the melanoma's photo in our dataset is taken from melanoma's dataset, which is available in Kaggle.

Allergic to the drug

Those who are allergic to the drug their photo is the most similar photo to the measles, which is almost impossible to identify between them. That is why we have chosen this. In addition, we have asked the authorities in that hospital, which is dedicated to the children, if there are any patients here that have skin diseases. They said we had last months but, unfortunately, now we don't have any patient who has skin diseases. Just there was some patient that had an anonymous disease and we have included it in a folder by the name of another in our collected dataset.

References

[1] Dr. Ananya Mandal, MD, Measles history, available: https://www.news-medical.net, Accessed: [1/August/20200]

[1]. World Health Organization, Measles-Afghanistan, available:https://www.who.int/emergencies/disease-outbreak-news/item/measles-afghanistan, Accessed: [10 February 2022]

[2]. Wikipedia Measles, available:https://en.wikipedia.org/wiki/Measles#:~:text=Measles%20is%20a%20highly%20contagious,runny%20nose%2C%20and%20inflamed%20eyes Accessed: [20,February 2020]

[4]. World Health Organization, Measles, available: https://www.who.int/news-room/fact-sheets/detail/measles?gclid=EAIaIQobChMIh9j_v_CI-AIVEJBoCR2cXgvUEAAYASAAEgJ5D_D_BwE [20,February 2020]

[5]. Healthline, Measles Symptoms, available: https://www.healthline.com/health/measles#symptoms [21,February 2020]

[6]. Center for Disease Control and Prevention, Measles (Rubeola), available: https://www.cdc.gov/measles/hcp/index.html [21,February 2020]

[7]. Helfand, R. F. et al. Diagnosis of measles with an IgM capture EIA: the optimal timing of specimen collection after rash onset. J. Infect. Dis. 175, 195–199 (1997).

[8]. Rota, P. A. et al. Improving global virologic surveillance for measles and rubella. J. Infect. Dis. 204, S506–S513 (2011).

[9]. Van Binnendijk, R. S. et al. Evaluation of serological and virological tests in the diagnosis of clinical and subclinical measles virus infections during an outbreak of measles in the Netherlands. J. Infect. Dis. 188, 898–903 (2003).

10]. Kimberly Glock1+, Charlie Napier1+, Andre Louie1, Todd Gary1, Joseph Gigante2, William Schaffner3, Qingguo Wang1*. Measles Rash Identification Using Residual Deep Convolutional Neural Network. 2021

[11]. Naqibullah Vakili, Jonathan H. Chan, Worarat Krathu, Nipat Phattarakijtham,

Kazuya Hirata. Hand Foot and Mouth Rash Detection Using Deep Convolution Neural Network. 2021. P.4

[12]. Nawal Soliman ALKolifi ALEnezi. A Method of Skin Disease Detection Using Image Processing and Machine Learning. Procedia Computer Science 163 (2019) 85–92. P.91

[13] Hubsche.J.M et al. Challenges of measles and rubella laboratory diagnostic in the era of elimination. April 2017, clinical Microbiology and infection 28(8), P.1

[14] Zarimina Islam, Tharwani.Zh, Butt.MS, Shaeen.S, Measles in Afghan Refugees: Challenges, Effort, and Recommendations, Clinical Epidemiology and Global health, February 2022, P.3

[15] Ahsan.M, Siddique.Z Machine learning passed disease diagnosis: A comprehensive review, [January,2,2022], P.3

[16] Wikipedia Online. Http://en.wikipedia.org/wik, accessed:[7 Aug , 2022]

[17] Essam H Houssein, Marwa M Emam, Abdelmgeid A Ali, and Ponnuthurai Nagaratnam Suganthan. Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. Expert Systems with Applications, 167:114161, 2021

[18] Virkramaditya Jakkula, Tutorial on Support Vector Machine (SWM), School of EECS, Washington State University Pullman 99164, P.10

[20] Burges C., "A tutorial on support vector machines for pattern recognition", In "Data Mining and Knowledge Discovery". Kluwer Academic Publishers, Boston, 1998, (Volume 2)

[21] Akkaya.B, Colakoglu.N, Comparison of Multi-class Classification Algorithms on Early Diagnosis of Heart Diseases, y-BIS 2019 Conference Book, September 2019, P.25

[22] Fan.Z et al, Image Classification Method Based on Improved KNN Algorithm, Journal of physics, School of Information and Technology, Shijiazhuang Tiedao University, Shijiazhuang, Hebei, China,2021, P.2

[23] Chollett.F, 2018, Deep Learning with Python, Manning Publications Co, 1st ed, United States of America, P.8

[24] Lan Goodfellow and yoshua Bengio and Aaron Courville, deep learning MIT press. 2016, P.326

[25] Manav Mandal, May 1-2021 Introduction to Convolutional Neural Networks (CNN), available: https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/, accessed: August-19-2021

[26] Convolutional Neural Networks, available: https://www.wikipedia.com/
Accessed: [19-August-2022]

[27]. Habibi, Aghdam, Hamed [2017-June-30]. Guide to convolutional neural networks: a practical application to traffic-sign detection and classification. Heravi, Elnaz Jahani. Cham, Switzerland. ISBN 9783319575490. OCLC 987790957 (https://www.worldcat.org/oclc/987790957)

[28] Shea.k, Nash.R, An Introduction to Convolutional Neural Networks, Department of Computer Science, Aberystwyth University, Ceredigion, SY23 3DB, [2-Dec-2015]

[29] EngEd Community, Getting started with image Preprocessing in Python, [August, 21, 2021] available: https//www.section.io/engineering-education/image-preprocessing-in-python/, accessed: [August, 20 2022]

[30] Oluwatobi. Ayodeji, et al, Preprocessing techniques Feature extraction, 2015, available: https://www.sciencediret.com/topics/computer-science/preprocessing-technique, accessed: [20 August 2022]

[31] Sultana.F et al, Evolution of Image Segmentation using Deep Convolutional Neural Network, Department of Computer Science University of Gour Banga India, [29 May 2020] P.3

[32] Sarki.R et al, Image Preprocessing in Classification and Identification of Diabetic Eye Diseases, Data Science and Engineering, [17, August 2021]. P.4

[33] Goswami.T et al, Skin Disease Classification from image – A Survey, 2020 6th, International Conference on Advanced Computing and Communication Systems (ICACCS), P.3