



Convolutional Neural Network-based medical checkup system for Pigmented Skin Lesions Classification.

School of Computing, Engineering and Mathematics
Coventry University

Bsc Computer Science
Project Supervisor: Dr.David Croft
Submission Date: 21/04/2020

Author Name: Vinayak Sareen
Student Id Number: 7651331
Ethics Application Number: P101878

Statement of originality

Declaration of Originality

I Declare that This project is all my own work and has not been copied in part or in whole from any other source except where duly acknowledged. As such, all use of previously published work (from books, journals, magazines, internet etc.) has been acknowledged by citation within the main report to an item in the References or Bibliography lists. I also agree that an electronic copy of this project may be stored and used for the purposes of plagiarism prevention and detection.

Statement of copyright

I acknowledge that the copyright of this project report and any product developed as part of the project, belong to Coventry University. Support, including funding, is available to commercialise products and services developed by staff and students. Any revenue that is generated is split with the inventor/s of the product or service. For further information please see www.coventry.ac.uk/ipr or contact ipr@coventry.ac.uk.

Statement of ethical engagement

I declare that a proposal for this project has been submitted to the Coventry University ethics monitoring website (<https://ethics.coventry.ac.uk/>) and that the application number is P101878.

Signature: .

Date: **07/03/20**

Abstract

The study is an attempt to use deep learning techniques in the classification of common pigmented skin lesions to detect skin cancer. Melanoma is serious skin cancer, and early diagnosis of such tumours can lead to the treatment of such skin problems at it's early stages. The objective of the research is to develop an artificial intelligence-based automated system for detection of pigmented skin lesions for people with less mobility. The study involves experiments with different architectures of convolutional neural networks and understanding the effects of hyper-parameters on developing deep learning model. The results obtained from tests performed during the development of intelligent models were compared with predicted resulted in lesions by medical professionals.

Contents

1	Introduction	6
1.1	Introduction to Problem	6
1.2	Objective of Research	7
1.3	Motivation and Targeted Audience	8
1.4	Perceptron Model	9
1.4.1	Inspiration from biological Structure of Neurons	9
1.4.2	Neural Structure	9
1.4.3	Perceptron Model	9
1.4.4	Mathematical Representation	10
1.5	Multi Layered Feed Forward Neural Network	11
1.5.1	Cost Function and Backpropagation Algorithm	12
2	Literature Review	14
2.0.1	Conventional Diagnosis Methods	14
2.0.2	Functioning of Convolutional Neural Networks	15
2.0.3	Lenet Model Architecture	16
2.0.4	SVM Algorithm to classify Pigmented Skin lesions	16
2.0.5	Border Detection Based System	17
2.0.6	Deep Feature to classify Pigmented Skin lesions	17
3	Methodology	18
3.1	Required Installation and Configuration	18
3.2	Conceptual Modelling	19
3.2.1	Sequance UML Diagram	19
3.3	Data Processing and Normalisation	19
3.3.1	Data Normalisation	20
3.3.2	Image Segmentation	21
3.3.3	Thresholding Segmentation Algorithm	21
3.4	Convolutional Model Training	23
3.5	Experiments for optimal Model optimiser	24
3.5.1	Adam Optimiser	24
3.5.2	SGD Optimiser	25
3.5.3	RMSProp Optimiser	25
3.5.4	Optimiser Accuracy Results	26

CONTENTS

3.6	Experiments with Hyperparameters	27
3.6.1	Increased convolutional layers	27
3.6.2	Learning Rate Experiment	28
3.6.3	Epochs Experiment	29
3.7	Segmented Model Training	30
3.8	Transfer Learning	31
3.9	VGG16 Architecture	32
3.9.1	Transfer Learning with VGG16	33
3.10	Lenet Architecture	34
3.11	Developing Web System	36
4	Evaluations and Results	37
4.1	Stastical Hypothesis Testing	37
4.2	Confusion Matrix	38
4.3	Most common method for diagnosing skin lesions	41
4.4	Automated System can help to patients to diagnose book prior appointments	41
4.5	Classification comparison with Participants	43
5	Discussion	63
5.1	Achievements of Research Project	63
5.2	Deficiencies of Research Project	63
5.3	Future Improvement	64
6	Project Evaluation	65
6.1	Supervisor Feedback	65
6.2	Legal and Ethical Issues	66
6.3	Research Challenges	66
6.4	Project Managment	67
6.4.1	Software Development Lifecycle	67
6.4.2	Kanban Work Managment System	68
6.4.3	Version Control System	68
7	Conclusion	69

Acknowledgements

I want to express my gratitude to my supervisor Dr David Croft for guiding me thought the investigations. The support of all medical professionals who contribute by providing data to the research is sincerely acknowledged. At last, I would like to especially thank my mother "Nidhi Sareen", who supported me finically and psychologically throughout this journey of computer science

Chapter 1

Introduction

1.1 Introduction to Problem

Skin cancer is categorized into two types: melanoma and non-melanoma skin cancer. Melanoma, is the most dangerous kind of skin cancer accounted for an estimated 16,000 deaths each year from 2014 to 2016 in the United Kingdom (Cancer Research UK, 2020). The melanoma tumor caused by melanocytes can result in uncontrolled and abnormal growth which can spread in the human body (Korotkov and Garcia, 2012). Detection and classification of unknown pigmented skin lesions can result in early diagnosis of the medical problem. The research data provided by a cancer research organisation in 2017 has shown that melanoma was the 20th most common disorder with new incidents of 81,00 and 83,00 in males and females respectively in the United Kingdom (Korotkov and Garcia, 2012). Dermoscopy is a non-invasive method of examining the pigmented skin, which includes microscopic imaging of the surface structure of pigmented skin lesions (Korotkov and Garcia, 2012). Early diagnosis of pigmented skin lesions is crucial to classify skin disorders to decrease mortality concerning particular skin disorders. Dermoscopy improves the detection of melanoma compared to the detection of disease with naked eyes by analysing the pigmented skin lesion. Previous studies have shown that such tumours can result in higher chances of better treatment and cure of disease by removing the tumour (Celebi et al., 2007). The current diagnosis method of detection involves using ABCD rule which considers the Asymmetry, Border irregularity, Colour irregularities, Darmascopic structures respectively of common pigmented skin lesions (Loescher et al., 2013). People working in busy work environments or less mobility can be victims of belated and slow diagnosis of such dangerous skin cancers. The automated analysis of pigmented skin lesions using artificial neural networks can be beneficial in the optical analysis of microscopic images of pigmented skin lesions. The primary targeted audience who benefits from the outcome is the people who are working in busy work environments or people with less mobility are best to use cases that can use such an automated

1.2. OBJECTIVE OF RESEARCH

system. Booking a prior appointment with medical professionals based on the urgency of detected medical problems can result in the immediate treatment of patients with more critical conditions. The people with less mobility such as older audiences or people with special needs can detect pigmented skin lesions through online systems in an inconvenient manner. Medical institutions can use such technologies to automate the process of pre-health checkups and overcome the problem of shortage of staff members in case of emergency. Such automated systems can also result in the faster diagnosis of medical problems compared to a manual analysis by a clinician. Furthermore, manufacturing companies that supply the microscopic medical instruments can also use such intelligent models with their products to provide value to customers and medical institutions.

1.2 Objective of Research

The research concentrates on developing a type of artificial neural network called a convolutional network to perform automated optimal analysis to identify the class of pigmented skin lesions. The research focuses on providing the quantitative analysis on comparing the results predicted by the automated intelligent machine are compared with medical professionals to identify the classes of pigmented skin lesions. The study employs different experiments by applying different model architectures and analysing accurate hyper-parameters for optimal performance. The research concentrate on analyzing limited skin tumors such as melanoma, benign keratosis, melanocytic nevi and basal cell carcinoma. The hypothesis for the research is that the model will perform in same time efficient as the medical professionals in examining the pigmented skin lesions. The investigation employs publicly available HAM10,000 dataset (Tschandl, 2018). The HAM10,000 dataset contains an extensive collection of 10,000 images of labelled data units that were collected from a diverse population of subjects over twenty years. The outcome of the research project will analyse the effectiveness of the automated pigmented lesion system. Furthermore, the intelligent model will be deployed on a web-based system to provide an interface to use the system to analyse by the general audience.

1.3 Motivation and Targeted Audience

Deep Learning technologies have always been fascinating for me to research on various models of neural networks. These artificial neural networks imitate the workings of the neural system and can be used as an intelligent system to recognise the visual patterns by analyzing the data. The primary motivation behind conducting the project is to research artificial neural networks to automate the optical classification tasks to help the community. The outcome of the research can be used by the general audience and healthcare institutions towards early diagnosis of common pigmented skin lesions. The secondary motivation behind developing such an automated deep learning system is to contribute towards reducing the mortality rates and improved pre-health checkup medical systems. An automated system can have an impact on the general audience by providing a convenient health checkup system which will detect melanoma and other common skin lesions. The primary targeted audience who benefits from the outcome is the people who are working in busy work environments or people with less mobility are best to use cases which can use such an automated system. Booking a prior appointment with medical professionals based on the urgency of detected medical problems can result in the immediate treatment of patients with more critical conditions. The people with less mobility such as older audiences or people with special needs can detect pigmented skin lesions through online systems in an inconvenient manner. Medical institutions can use such technologies to automate the process of pre-health checkups and overcome the problem of shortage of staff members in case of emergency. Such automated systems can also result in faster diagnosis of medical problems compared to a manual analysis by a clinician. Furthermore, manufacturing companies which supply the microscopic medical instruments can also use such intelligent models with their products to provide value to customers and medical institutions.

1.4 Perceptron Model

1.4.1 Inspiration from biological Structure of Neurons

The human brain is composed of millions of specialised cell "neurons" which are interconnected to each other, which carry electrical and chemical signals from neuron to another to function. There are an estimated 500 trillion connections between neurons in the human nervous system, which helps communicate signals (Patterson and Gibson, 2017). The inspiration for artificial neural networks is taken from the functioning of the human brain and learn to recognise patterns and relationships in the data (Agatonovic-Kustrin and Beresford, 2000).

1.4.2 Neural Structure

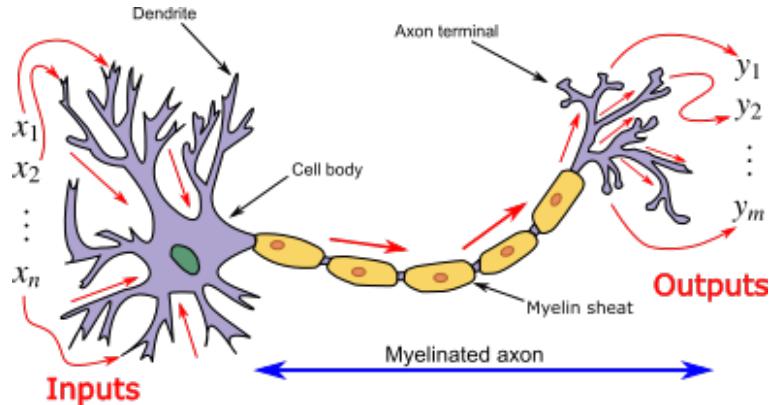


Figure 1.1: Human neural structure

Figure 1.1 represents the biological structure of neurons that helps in the communication of electric signals in the nervous system to learn and process information. The biological structure of neurons includes three significant parts dendrites which are responsible for accepting the electrical and chemical signals to the neuron. Furthermore, the neuron contains the nucleus, which is accountable for the processing input information with the neurons. At last, the processed information is passed to another neuron which is interconnected to neurons in the human nervous system through axon terminals (Agatonovic-Kustrin and Beresford, 2000).

1.4.3 Perceptron Model

Perceptron model was proposed by Frank Rosenblatt in 1943 to design the model to mimic the human brain (Kussul et al., 2001). Perceptron model or the single-layered feed forwards network networks take the vectors of inputs and multiply with a randomly initialized weights and add random bias to network

1.4. PERCEPTRON MODEL

and process the information by providing data to the activation function to process the information (Agatonovic-Kustrin and Beresford, 2000).

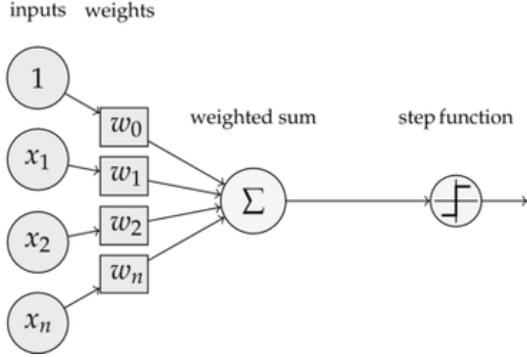


Figure 1.2: Perceptron Model

The figure 1.2 represents the simple perceptron model which consists of inputs x and weights w for which the weighted sum of multiplication result of inputs and weights will be passed to the step activation function.

1.4.4 Mathematical Representation

The equation below represents the preceptron model mentioned above in Mathematical notations.

$$y = \left[\sigma \left(\sum_{k=0}^n x_k \cdot w_k + b_k \right) \right] \quad (1.1)$$

In, equation above $\sigma = Activation\ function$, x and w represents the inputs provided and weights vectors in the network respectively. Furthermore, the symbol b represents the bias in the equation, during the optimization of the network weights and bias are adjusted accordingly to improve the prediction. There are various activation functions σ such as step function sigmoid, relu, leaky relu and others that can be applied based on the requirements of the model prediction.

1.5 Multi Layered Feed Forward Neural Network

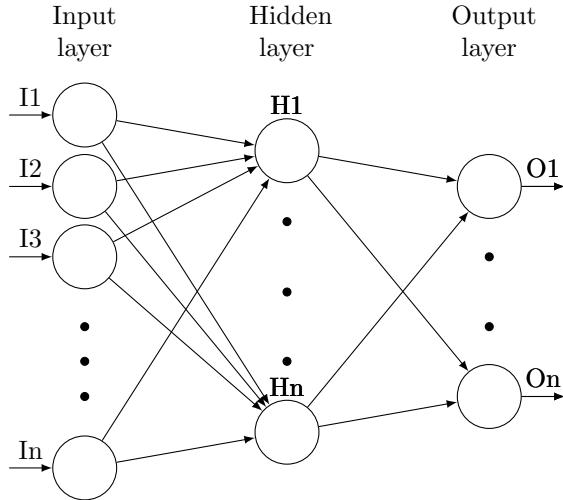


Figure 1.3: Multi Layered Neural Network diagram.

The perceptron model acts as the base for all the functioning of modern multi-layered neural networks. The output of the single neuron described in the above perceptron is taken as the input in the next layer of the network. The above figure(1.1) shows an example of multi-layered neural where the first layer is known as the input layer and intermediate layers are called hidden layers and the last layer is known as the output layer. The single neuron can perform limited computation but the computation power increases with inter-connected neurons in the network (Agatonovic-Kustrin and Beresford, 2000) where at each layer information is processed through the appropriate activation function. Artificial neural networks are designed to learn the patterns and relationships in the data and requires enough amount of data to train and predict accurate outputs (Agatonovic-Kustrin and Beresford, 2000). The training data is passed to the neural network to recognize the patterns in data and each iteration or epoch the model prediction gets improved with the optimisation algorithm which propagates through the network to update the weights and bias to increase the accuracy (Agatonovic-Kustrin and Beresford, 2000). The accuracy of the neural networks are determined through various hyper-parameters such as learning rate, number of hidden layers in the model and number of epochs for which the model is trained.

1.5.1 Cost Function and Backpropagation Algorithm

The objective for training deep learning models is to reduce the error in prediction on each iteration and learn data patterns and relationships by increasing the prediction accuracy. Cost function is an appropriate method to analyse the performance of the neural network as it describes how well, model is predicting the actual results. The equation below describes the cost function equation also known as mean squared error (Lubis et al., 2014).

$$c = \frac{1}{2m} \sum_{i=1}^m (h - y)^2 \quad (1.2)$$

In, equation(1.2) h is the predicted outcome of the intelligent model and y is the actual value of the input. The equation above is mean square of predicted value - actual value. The objective of optimisation algorithms is minimise the equation(1.2) to decrease the value of cost function so, the value is close to zero as much as possible so, that predicted and actual value are almost same which means overall model accuracy has improved (Lubis et al., 2014).

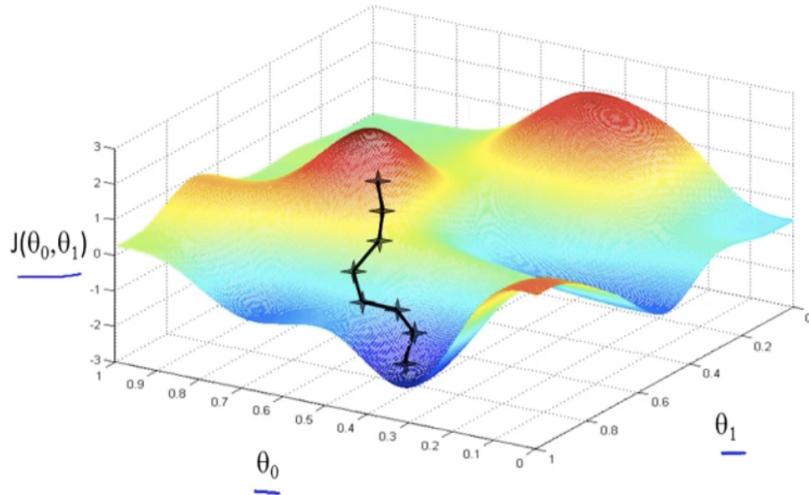


Figure 1.4: Gradient Descent

Gradient descent algorithm is a optimisation algorithm which is iterative in nature aims to minimise the cost function is gradient descent. The algorithm functions by minimising the objective function in the space.

$$\frac{\partial}{\partial w} \left(\frac{1}{2m} \sum_{i=1}^m (h - y)^2 \right) \quad (1.3)$$

$$\frac{\partial}{\partial b} \left(\frac{1}{2m} \sum_{i=1}^m (h - y)^2 \right) \quad (1.4)$$

The algorithm functions by finding the partial derivative of the cost function in respect to weights and bias as shown in the equations (1.3) and (1.4) above. The objective is to find minimum of the cost function which improves the accuracy (Lubis et al., 2014). Backpropagation utilises the gradient descent algorithm and propagates after each layer to update the bias and weights in backwards direction to improve accuracy of the model.

Chapter 2

Literature Review

2.0.1 Conventional Diagnosis Methods

The most common conventional diagnosis method of detection involves using ABCD rule which considers the Asymmetry, Border irregularity, Colour irregularities and Dermoscopic structures respectively of the common pigmented skin lesions (Loescher et al., 2013). The rule for classification was introduced in 1985 as abcd rule and was amended with abcde in 2004 where ‘E’ reflects the lesions which are evolving. An alternative method of examination of pigmented skin lesions is ugly duckling signs which was introduced to state the limitation of the ABCD rule (Daniel Jensen and Elewski, 2015). The ugly duckling signs state the spot which is unlike other lesions are great suspects of melanoma (Grob and Bonerandi, 1998). Despite the limitations of above methods, both provide a great framework for clinicians and general audiences to spot melanoma based symptoms (Daniel Jensen and Elewski, 2015). Small malignant melanoma are millimeters in size in initial phases of its growth.

The micro-melanoma requires more attention to be detected by physicians and researchers believe that small lesions possess challenges for medical professionals to clearly examine malignant based problems (Bono et al., 2004). Dermatoscopy is non-invasive microscopic imaging of pigmented skin lesions which provides clear imaging to perform proper analysis on pigmented skin lesions (Loescher et al., 2013). The study was conducted in Italy for the period of five years in which ninety four melanoma based lesions. Furthermore, results were treated to examine the clinical and dermoscopic features of cutaneous melanoma lesions with maximum diameter of 3mm. The outcome of the research has shown that only samples of twenty two lesions which accounts for 2.4 percent of overall samples were decided based on the clinical size feature of pigmented skin lesions (Bono et al., 2004). Moreover, the research mentions that dermoscopic results are more accurate to diagnose and has been an aid for early detection of melanoma skin cancer provided the clinicians are aware about disease (Bono et al., 2004). The result of dermatoscopic images is examined by dermatologists

to classify the pigmented skin lesion. Dermatologists select different approaches to examine pigmented skin lesions which might include ABCDE rule or ugly duckling signs, where both the methods complement each other to detect pigmented skin lesions. The research proposed an addition of ‘F’ to mnemonic which uses both current methods ugly duckling sign and ABCDE rule which also account for funny or unlike lesions to be suspect of melanoma based lesion (Bono et al., 2004).

2.0.2 Functioning of Convolutional Neural Networks

The convolutional neural networks are the feed-forwards neural networks which contains alternative convolutional layers and sampling layer which at the end is connected to fully connected neural networks (Liew et al., 2016). The objective of the convolutional layers is to extract the features from the given input image samples by applying filters which is used in the training of overall network.

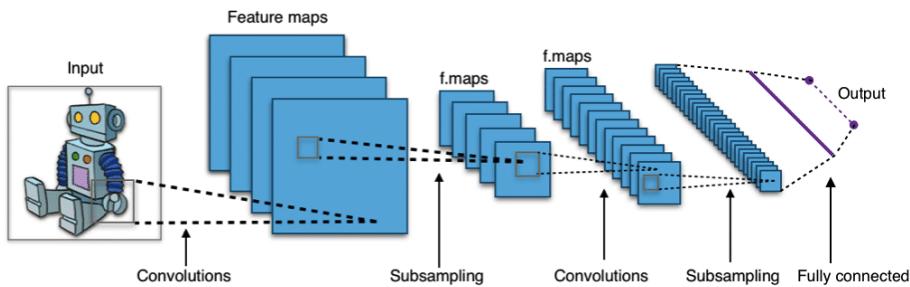
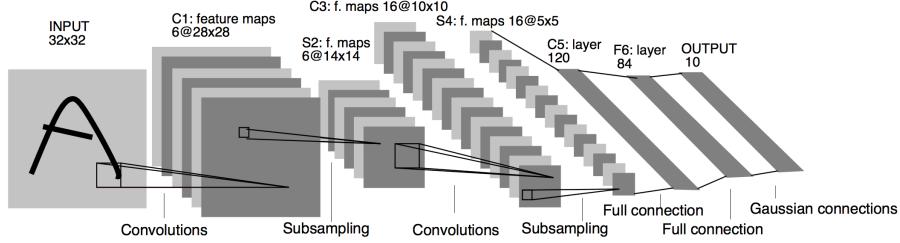


Figure 2.1: Convolutional Neural Network

The initial layer is known as the input layer in convolutional network which contains the pixel information of the image. The convolutional layer of the network determines the output of the neuron of particular region by applying image kernels to extract the features from the images (O’Shea and Nash, 2015). The image kernels are the matrix of filter which are computed with the local region and produce the feature maps (O’Shea and Nash, 2015). Furthermore as shown in 2.1 sub sampling also known as downsampling is applied to the feature maps to reduce the dimensionality of the feature maps. In addition, there can exist various number of convolutional layers and sub sampling layer based on the architecture of the network. The next layer in the network is flattening layer which reshapes the feature maps extract from previous convolutional layers of the network. At last the flattened information is passed to the fully connected neural network to analyse the relationships in the data and make appropriate predictions (O’Shea and Nash, 2015).

2.0.3 Lenet Model Architecture



In 1998, the lenet-5 the model architecture was purposed by Yann Lecun in an attempt to recognise handwritten images of mathematical digits. The figure 2.0.3 shows the lenet-5 model architecture, the model only consists of seven overall layers and two layers are convolutional to extract the feature maps from given $32 * 32$ pixel of the image (LeCun et al., 1998). Lenet architecture focuses on applying the $5 * 5$ convolution to extract 6 feature maps of size $28 * 28$ pixels. Furthermore, sub-sampling is performed on the feature maps using the average pooling algorithm and contains the 6 feature maps of size $14 * 14$ followed by a convolutional layer of 16 feature maps of size $10 * 10$ (LeCun et al., 1998). Further process of downsampling is performed on feature maps using average pooling and processed through a fully connected neural network. The model was used in the banks to detect handwritten digits on the documents. The lenet architecture was one of the earliest convolutional network to perform the automation in banking sector. The research project will focus on using the lenet convolutional model architecture on detecting the pigmented skin lesion to investigate it's performance.

2.0.4 SVM Algorithm to classify Pigmented Skin lesions

Thompson Felsia and Jeyakumar proposed research in 2017 on support vector machine based classifier to detect multi-lesions skin cancer by analysing pigmented skin lesions with an accuracy of 86.37 percent. The proposed investigation with SVM based classifier has performed image segmentation using SRM (support region merging) algorithm. Furthermore, it employs SURF (speed up robust features) to find the region of interest for feature extraction to get optimal classification performance based on vector-based technique (Thompson and Jeyakumar, 2017). However, the research does not include image augmentation which generalises the predictions accurate to test in real-world environment. The research papers mention that support vector machine for automated classification of pigmented skin lesions is sensitive to the artefacts and can potentially increase the false positives which mean that predicted result for analysis was wrong positive prediction instead of an actual negative result. The investigation will perform image augmentation to generate random samples of images with different

rotation angle and flipped images will be used to train and test the model to generalise the overall performance.

2.0.5 Border Detection Based System

Rahil Garnavi and his other co-researchers purposed research based on a state of the art border detection method combined with the colour space analysis and clustering-based histogram hybrid thresholding to classify pigmented skin lesions. The research was primarily focused on the research was to develop the hair removal mechanism to perform colour channels transformation. Furthermore, for all the image channels the noise reduction and clustering-based histogram thresholding were performed for optimal border detection. The predicted outcomes of novel border detection system were compared with the borders detected by the actual dermatologists on a sample of dermoscopic pigmented skin lesions to understand the reliability of the system (Garnavi et al., 2011). However, the system was only tested on a data sample set of 30 dermoscopic images and four sample sets of dermatologist hand-drawn images were used as ground truth to compare the results. The system was tested on overall 85 dermoscopic pigmented skin images with high resolution. Border detection can be used to analyse the pigmented skin lesions but convolutional networks have the potential to find more data patterns in the images to minimise the cost function using the backpropagation algorithm. The current research will employ basic image segmentation based on the binary threshold algorithm as an experiment to help network detecting more accurate borders of pigmented skin lesions.

2.0.6 Deep Feature to classify Pigmented Skin lesions

In 2016, a research paper from Simon Fraser University's computer science and medical image analysis lab had researched using deep residual network architecture with ten labelled classes of pigmented skin lesions. The research was based on very deep convolutional network architecture with the accuracy of 85.8 percent in classifying five distinct classes and 81 percent in classifying 10 classes of pigmented skin lesions (Kawahara et al., 2016). Although the performance of the overall convolutional network was accurate, the training and testing data were limited to 13,00 overall images of 10 distinct classes. However, In the current research project, the classes of labelled images will be five and around 9,000 overall images will be used during the investigation. Estimated 80 percent of data will be consumed for training the model, and the rest of the label images will be used as validation and testing datasets to evaluate the performance of the model. Research is also consuming such artificial neural network-based technologies to various areas of investigations.

Chapter 3

Methodology

3.1 Required Installation and Configuration

The research purposes a solution based on deep convolutional neural networks With various experiments mentioned in further sections. The experiments require specific configuration for cuda libraries to run programs on available GPU(Graphical Processing Unit) for faster processing and further, instructions to setup environment can be found at Tensorflow gpu installation guide ¹. Alternatively, the machine learning models can also run without installing cuda libraries on CPU but the processing will be slow. In addition, the convolutional neural network was implemented using Python3 and Jupyter notebooks were used in these experiments which provides an appropriate interface to experiment and write markdowns. Jupyter notebooks can be installed by installing Anaconda ². Furthermore, data science and machine learning libraries are required such as Numpy, Keras, Matplotlib and OpenCV. Numpy library is required for performing mathematical operations on multi-dimensional NumPy arrays. The matplotlib library provides an interface to visualise the output results. The OpenCV library was used for image processing in the process of developing an automated system for classifying pigmented skin lesions. Moreover, keras library was used to develop deep convolutional models. HAM,1000 (Human Against Machine, 1000) dataset was used to develop image classifier (Tschandl, 2018). The dataset primarily contains two folders which includes overall 10,000 dermatoscopic images of pigmented skin lesion. In addition, it also contains a csv file which includes the meta-information regarding pigmented skin lesions. The required files can be downloaded from harvard dataverse webiste ³.

¹<https://www.tensorflow.org/install/gpu>

²<https://www.anaconda.com/distribution/>

³<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T>

3.2. CONCEPTUAL MODELLING

3.2 Conceptual Modelling

The Conceptual models of the system was designed using workflow and UML(unified modelling language) diagrams to visualise behavioural requirements of the automated system. The uml diagram was produced using visual paradigm software which provides interface to produce digital architecture and system design of software solution.¹

3.2.1 Sequence UML Diagram

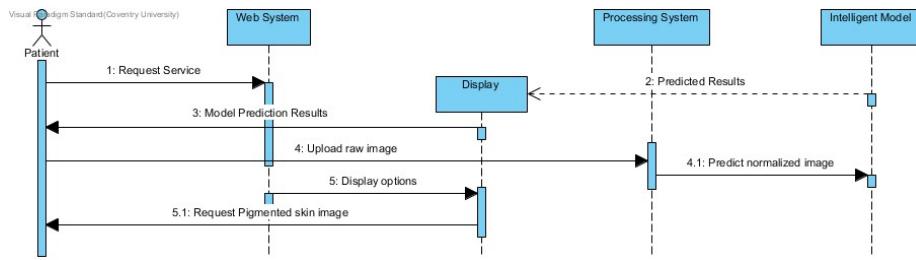


Figure 3.1: Sequence Diagram

The figure 3.1 represents the sequence uml (unified modelling language) diagram of the system. The initial actor in the system is the user or the patient of the pigmented skin lesion. The diagram shows the flow of the messages among different lifelines in the system. The messages between lifelines outlines the interaction of the patient with the system in sequential fashion. Therefore, the diagram helps to capture the behavioural requirements of the automated system.

3.3 Data Processing and Normalisation

The information was read using pandas into the data frame, which is a data structure that allows storing tabular data from CSV files as shown in figure 3.2 which represents the small section of the whole pandas dataframe. The CSV file contains irrelevant information such as age, sex and localisation of patients in the data frame which was removed by dropping the non essential columns. In addition, the dataset contains unclear and hairy images of pigmented skin lesions which were manually removed from the dataset to enhance the quality of available data. Furthermore, the research only focuses on limited categories of pigmented skins which results in dropping data columns for the other categories of data. The dataframe contains a lesionid column which corresponds to image names which were read into numpy array using pillow library from respective directories. The data for convolutional neural networks needs to divided into

¹<https://www.visual-paradigm.com/>

3.3. DATA PROCESSING AND NORMALISATION

	lesion_id	image_id	dx	dx_type	age	sex	localization
0	HAM_0000118	ISIC_0027419	bkl	histo	80.0	male	scalp
1	HAM_0000118	ISIC_0025030	bkl	histo	80.0	male	scalp
2	HAM_0002730	ISIC_0026769	bkl	histo	80.0	male	scalp
3	HAM_0002730	ISIC_0025661	bkl	histo	80.0	male	scalp
4	HAM_0001466	ISIC_0031633	bkl	histo	75.0	male	ear

Figure 3.2: PSL Information Dataframe

training and testing data, the model learnings are performed with learning the patterns and relationships in the training datasets and evaluation of the model is performed on the testing data which is never feed to the intelligent model which training. The dataset was divided into training and testing sets using `sklearn.model_selection.train_test_split` class in the portion of 80 per cent for the training dataset and 20 per cent of testing datasets. The next step towards preparing the dataset was reading the images data into NumPy array for both training and testing datasets and converting the image names from pandas series to NumPy array corresponding to each image and assign class number based on category of pigmented skin lesion in the dataset. Furthermore, the training and testing datasets were serialised into dictionary in a pickle encoded file. Therefore, the encoded file sizes are compact and are portable in comparison to storing actual image files.²

3.3.1 Data Normalisation

The images with RGB(Red, Green and Blue) channels information was stored in NumPy multi-dimensional array with numbers ranging from 0 to 255. The NumPy array was converted into the float32 format, and each element of the array was divided by 255 to normalise the data so, that it only ranges between 0 and 1 in float format which will help while training the model. In addition, one hot encoding was performed on class labels of the pigmented lesions. The one-hot encoding is a representation of the categorical variable as binary vector and normalises the categorical labels into a binary vector. One hot encoding was performed to the labels data for each pigmented skin lesions. The data normalisation process will help to train the models as images from training and testing data samples are ranging from 0 to 1 instead of 255 which results in the improved model accuracy.

²<https://github.coventry.ac.uk/sareenv/Final-Year-Project/blob/master/Research>

3.3. DATA PROCESSING AND NORMALISATION

3.3.2 Image Segmentation



Figure 3.3: Pigmented Skin lesions before Image Segmentation

Figure 3.3 shows the sample from training dataset before performing image segmentation. The figure 3.3 above was plotted using matplotlib library.

3.3.3 Thresholding Segmentation Algorithm

Thresholding is one of the commonly adopted method in image segmentation which helps in discrimination most significant pixels in the images (Al-Amri et al., 2010). The threshold value is selected and the gray scale images are converted into the binary representation of the image and value of image which are greater than the threshold value will be selected with keeping all the attributes of the images such as position and shape (Al-Amri et al., 2010). Thus, reducing the complexity of the image data and making it easier for classification related tasks. Furthermore, the segmented images will be consumed in the model training. The thresholding segmentation was performed using OpenCV library using `cv.threshold(image, 0.5, 1, cv.THRESH_BINARY)` where threshold value of 0.5

3.3. DATA PROCESSING AND NORMALISATION

and maximum value of the pixel can be 1 as it was normalised.

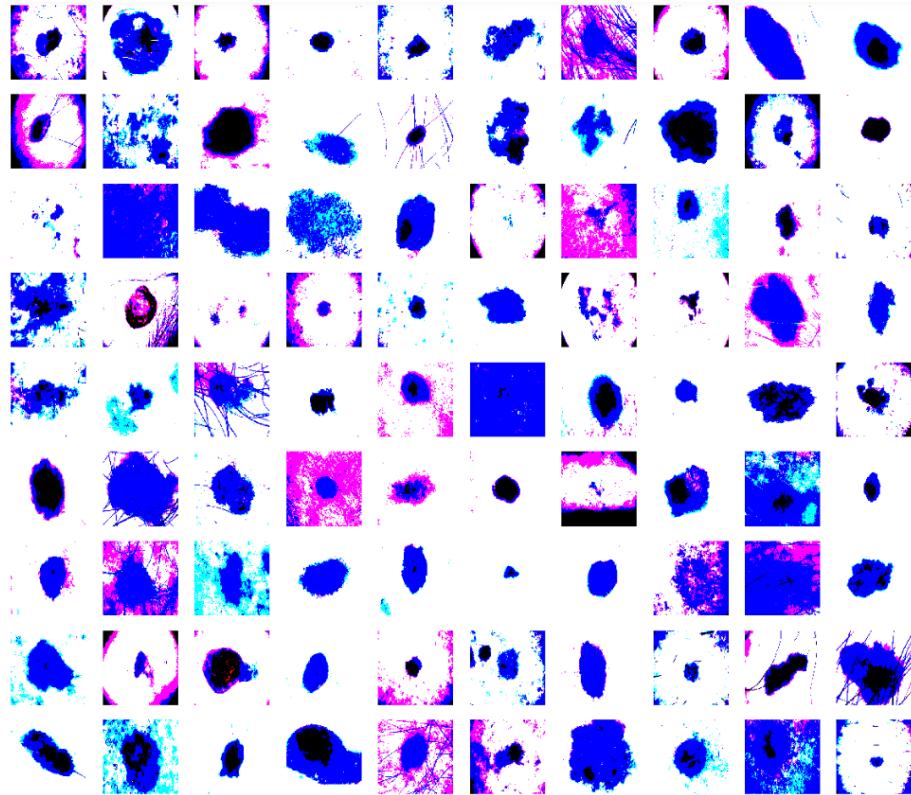


Figure 3.4: Binary Images Segmented Pigmented Skin lesions

The figure 3.4 shows the result of applying the threshold image segmentation on pigmented skin lesions. The process of image segmentation will simplify the process of border and edge detection by the convolutional neural network. Figure 3.4 was plotted using matplotlib library.
³

³<https://github.coventry.ac.uk/sareenv/Final-Year-Project/blob/master/Research>

3.4 Convolutional Model Training

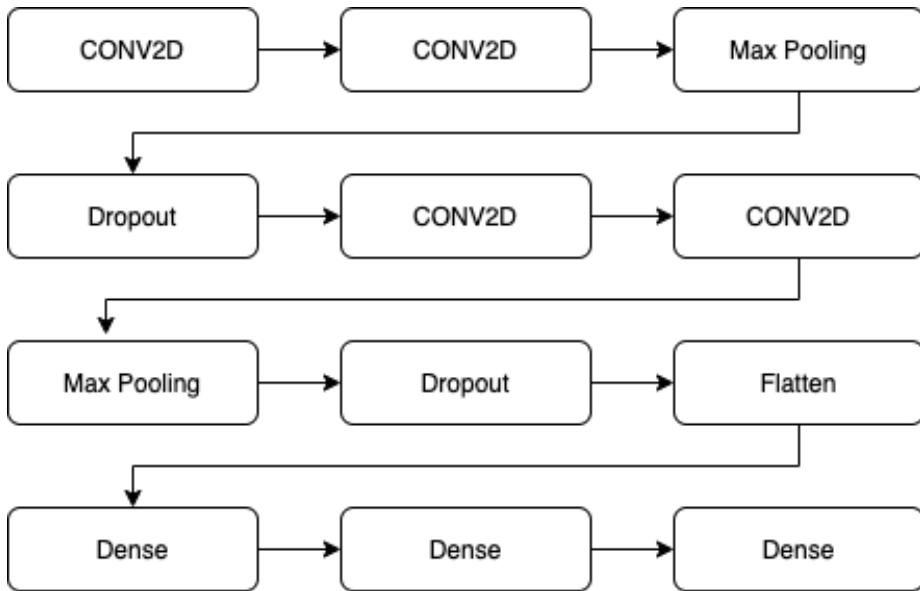


Figure 3.5: Model Architecture 1

The model architecture 3.5 was implemented using the keras library in which 2 dimensional convolutional layers were added to the sequential network, where width and height of input image is 224 pixels and the three rgb channels depth are represented by 3. The initial convolutional layers contains 32 input filters with the kernel size of (3, 3) with relu activation function. In the network layers followed by first two convolutional layers are polling layer in the architecture MaxPooling was used to extract maximum of the input features after applying image filter or kernel to the given image of pigmented skin lesions. Furthermore, dropout of 0.4 was used in the network to generalise the overall performance of the network and avoid overfitting of data points. The next two layers in the networks are also another convolutional layers with 64 image filters and similar relu activation function. Similar fashion as above was applied to the network with MaxPooling to extract most significant pixels from feature maps followed by the dropout in the network to generalise it. The features extracted by the convolutional layers are flattened into one dimensional array. The flattened array will be passed to the fully connected layers in the neural network to process the information. The model contains three dense hidden layers and one output layer in the neural network. Furthermore, the model architecture was compiled using various hyper-parameters which effects such as learning rate and optimiser for the convolutional model which helps in computing the gradient for the loss function to minimise the error in predicting category of pigmented skin lesion.

3.5 Experiments for optimal Model optimiser

The model architecture at initial stage was compiled using different hyper-parameters such as optimal optimiser for the neural network and learning rate. These hyper-parameters can have impact in finding gradient descent of loss function. Keras libraries provides various optimisers such as Adam, SGD and RMSProp Optimiser which aims towards reducing the cost function. The model architecture mentioned above was trained for 30 epochs or iteration to investigate the optimal optimiser for classification of pigmented skin lesions. The model was trained under learning constant rate of 0.001 and loss function of categorical crossentropy because the model has to evaluate the mutiple data classes for pigmented skin lesions.

3.5.1 Adam Optimiser

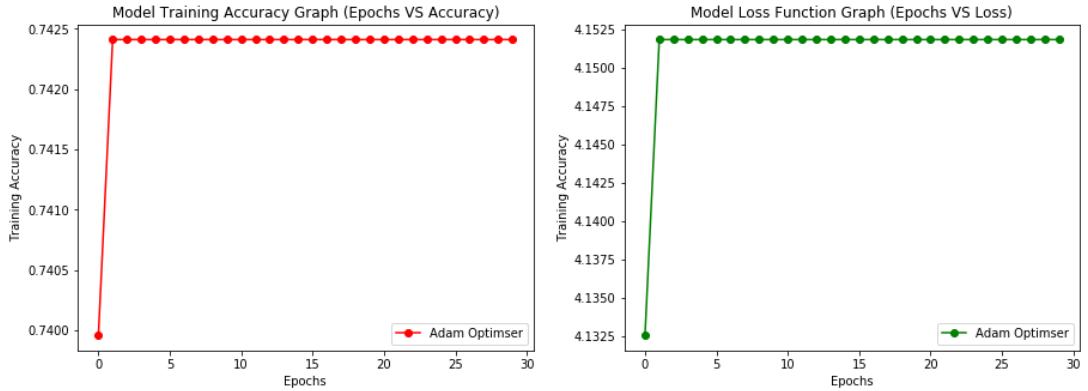


Figure 3.6: Model Results obtained with Adam Optimiser

The figure 3.6 refects the results obtained from training the model using adam optimiser. The figure 3.6 shows that model loss function is not decreasing and after few epochs there was no improvmant in model accuracy which means that adam optimiser is suitable optimiser for model architecture 3.5 presented above.

3.5. EXPERIMENTS FOR OPTIMAL MODEL OPTIMISER

3.5.2 SGD Optimiser

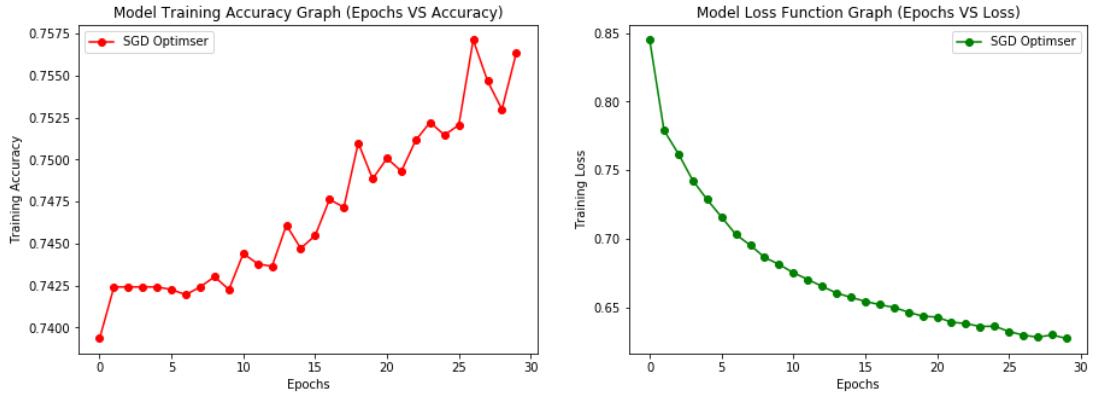


Figure 3.7: Model Results obtained with SGD Optimiser

The figure 3.7 reflects the results obtained from training the model using sgd optimiser. The model shows improvement in model performance. The training accuracy of model increased over epochs and loss function was decreasing as anticipated as shown in the figure 3.7 above.

3.5.3 RMSProp Optimiser

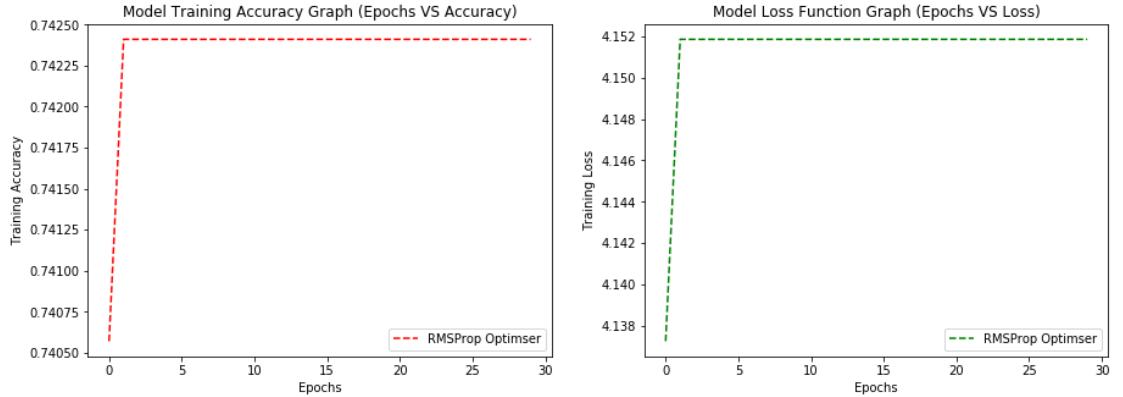


Figure 3.8: Model Results obtained with RMSProp Optimiser

The figure 3.8 reflects the results obtained from training the model using 'RMSProp' optimiser. The model performance was similar to the adam optimiser. Therefore, the SGD optimiser was optimal optimiser for classificaiton model under the

3.5. EXPERIMENTS FOR OPTIMAL MODEL OPTIMISER

constant hyper parameters mentioned above. All further experiments are performed on convolutional neural network are trained on SGD optimiser.

3.5.4 Optimiser Accuracy Results

Optimiser	Learning Rate	Model Accuracy	Epochs
ADAM	0.001	72.79%	30
RMSProp	0.001	72.79%	30
SGD	0.001	75.05%	30

The table above shows the accuracy results obtained from evaluating the trained model on testing data which were trained for 30 epochs or iterations with different optimiser.

⁴

⁴<https://github.coventry.ac.uk/sareenv/Final-Year-Project/tree/master/Research>

3.6 Experiments with Hyperparameters

The experiments were performed on the convolutional model to understand the effects of the hyper-parameters such as learning rate of the network, the number of epochs for which the model was trained and increasing the hidden and convolutional layers. The adjustments to the hyperparameters which were made on the model had an impact on the accuracy of the overall classification of the pigmented skin lesions.

3.6.1 Increased convolutional layers

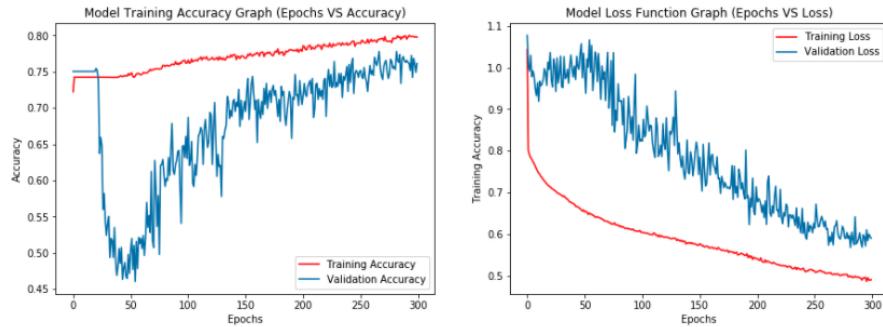


Figure 3.9: Convolutional Neural Network

The model architecture was changed with increase of four more convolutional layers in the network to detect more feature maps from images. The additional convolutional layers contain filters of 128 and 256 respectively. Furthermore, the model was trained for three hundred epochs with validation data to adjust the weights and improve the accuracy of the model with learning rate of 0.001 using SGD optimiser from the above experiments. The figure 3.9 shows the graph of increase in model accuracy of training and validation data over three hundred epochs. Furthermore, the diagram also shows the decline in the loss or cost function of the model. The model was evaluated on the testing data and resulted in 77.4 % accuracy in detecting the pigmented skin lesions.⁵

⁵<https://github.coventry.ac.uk/sareenv/Final-Year-Project/blob/master/Research>

3.6. EXPERIMENTS WITH HYPERPARAMETERS

Model Training without validation data

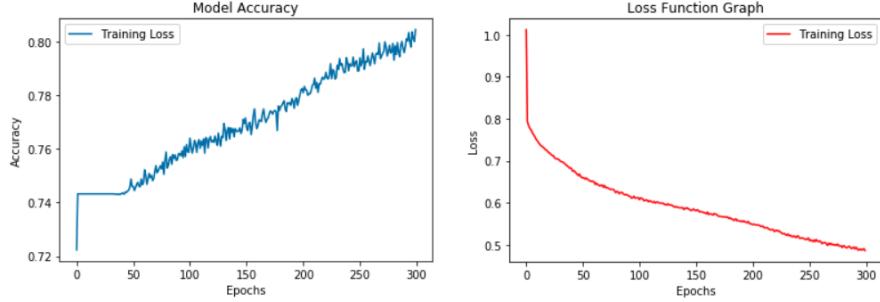


Figure 3.10: Model training without validation data

The experiment was performed with the same model architecture as above without providing the validation data to the model. The figure 3.10 show the increase in the model accuracy and decline in the the loss function over three hundred epochs. The accuracy of the model was evaluated to be 74.39% on testing data. Therefore, the further model experiments were performed with validation data.

3.6.2 Learning Rate Experiment

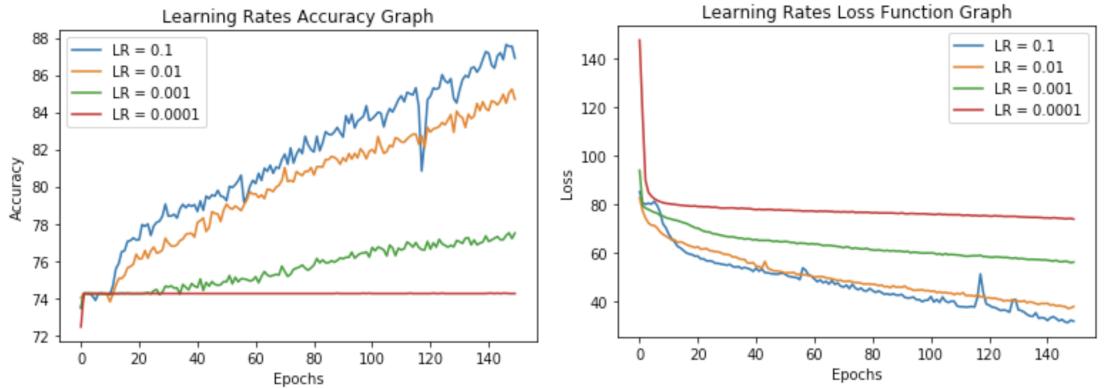


Figure 3.11: Variable Learning Rates

The figure 3.11 above shows that the model was trained for different learning rates as shown in the legend of the figure 3.11 for thirty epochs or iterations.

3.6. EXPERIMENTS WITH HYPERPARAMETERS

The outcome of the above test was that the model accuracy of the model were increasing with decrease in the learning rates. The graph obtained shows that the model accuracy was proportional to the learning rate while using SGD optimiser. The figure 3.11 also shows the decline in the loss function for different learning rates in convolutional model trained with SGD optimiser. Therefore, the most optimal learning rate to train the convolutional network neural network was 0.1 and 0.001. During the experiments, it was observed that model with the lower the learning rate consumes more time in the process of training. Further model training experiments are performed on learning rate of 0.01 to achieve the accuracy and relative speed to train the model.

Learning Rates Accuracy Results

Training Time	Learning Rate	Test Accuracy	Epochs	Optimiser
02:31:15	0.1	73.3%	140	SGD
02:32:52	0.01	81.01%	140	SGD
02:33:00	0.001	69.75%	140	SGD
02:33:22	0.0001	72.7%	140	SGD
02:33:39	0.00001	72.7%	140	SGD

The results presented above were evaluated on testing data and shows the direct relation of the time required to train convolutional neural network where decreasing learning rates requires more time to train models. These results are obtained by training the model on Nvidia GTX-1080 gpu(graphical processing unit), training time might differ on different gpu based on its computational power.

3.6.3 Epochs Experiment

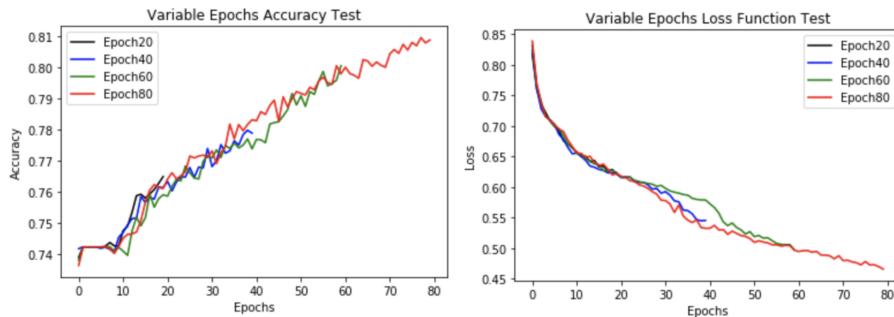


Figure 3.12: Results obtained for different epochs

The figure 3.12 shows the accuracy of model with same architecture for different number of epochs. The model accuracy is directly proportional to the

3.7. SEGMENTED MODEL TRAINING

number of epochs as training the neural networks is optimisation problem and objective is to find minimum of the cost function. The model accuracy on training data improves over each epoch as the model finds local minimum of cost function at each epoch and improve the prediction in the next iteration. However, when the model reaches the global minimum of the objective function there will be improvements in model accuracy. The figure 3.12 shows the model trained with the most number of epochs which is eighty in these experiments has least value of cost function and maximum training accuracy.

Learning Rate	Test Accuracy	Epochs	Optimiser
0.01	73.61%	20	SGD
0.01	76.97%	40	SGD
0.01	79.46%	60	SGD
0.01	78.68%	80	SGD

The table above shows the results obtained from evaluating model accuracy on the test data trained over different number of epochs. The general trend can be observed that with the increase in the epochs model accuracy was also observed to be improved.

3.7 Segmented Model Training

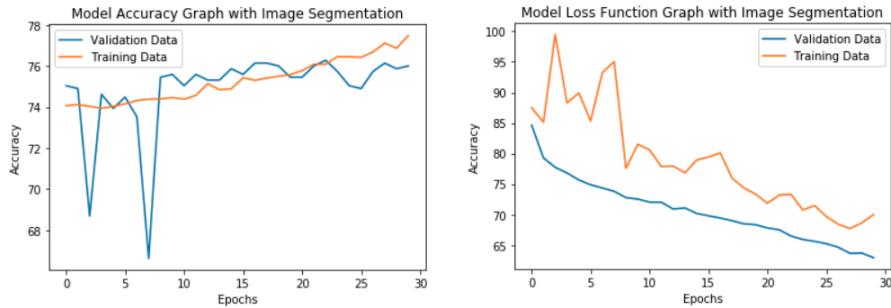


Figure 3.13: Segmented Model Training

The model training was performed using the segmented images of pigmented skin lesions as mentioned in the image processing section. The model with same architecture was trained under thirty epochs using SGD optimiser. The experiment performed above resulted in 74.0% model accuracy on testing data. The figure 3.13 shows the increase in accuracy rate of training and validation data and decline in the loss function.

3.8 Transfer Learning

Transfer learning is a method to consume network weights from previously trained model and applying the model weights in training of new convolutional neural network. The objective of the experiment was to consume the weights obtained from training of model with segmented data and use it in the processing of training same model architecture using normal pigmented skin lesions.

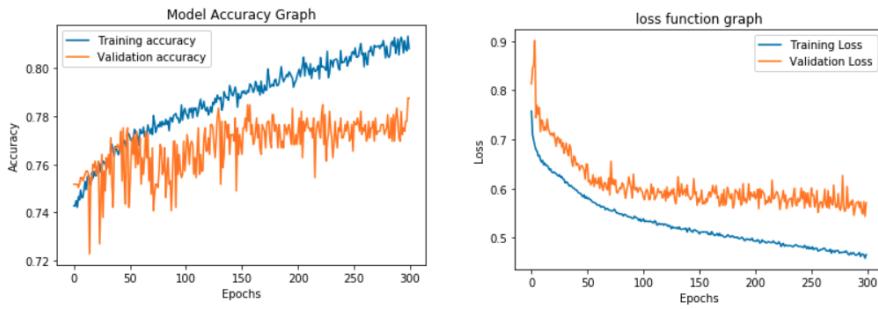


Figure 3.14: Transfer Learning Model

The figure 3.14 shows the graph indicating increasing in the model accuracy for training and validation data and decline in the loss functoin of the model. The model was trained using the SGD optimiser for 300 epochs and had produced the accuracy of 79.02% on the evaluation data. The result of applying the transfer learning can be observed as without applying the weights from segmented model training the accuracy was evaluated to be 77.4%. Further experiments are performed using the transfer learning to evaluate the impact on model performance.

3.9 VGG16 Architecture

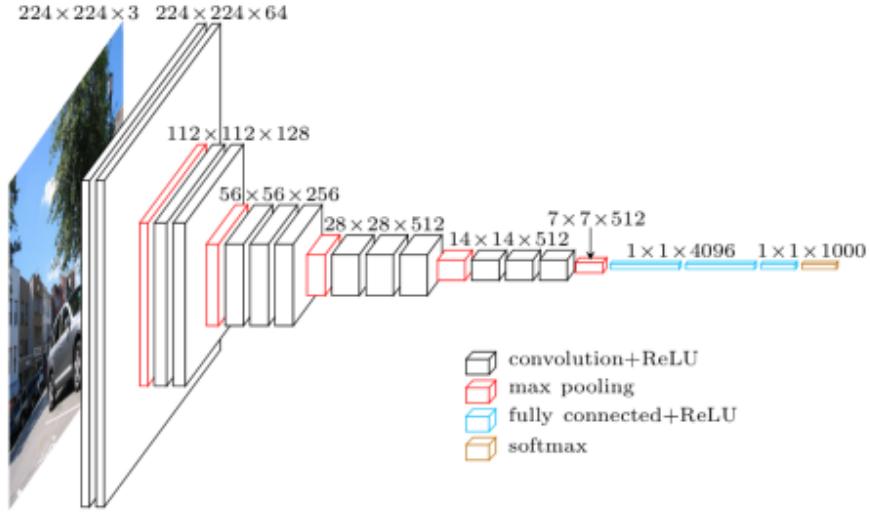


Figure 3.15: VGG16 Model Architecture

The VGG16 model is heavy model architecture containing 16 layers in the model. The model architecture consist of 13 convolutional, 5 pooling and 3 fully connected layers. Due to the heavy model architecture with estimated 138 million parameters and limited hardware resources it was not possible to trian the model on personal machine. As, a result the google colab platform was used develop the convolutional network ⁶.

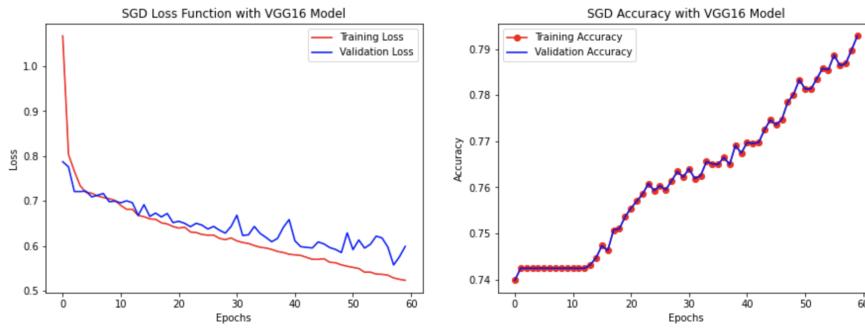


Figure 3.16: VGG16 Model Architecture

The figure 3.16 shows the graph of model accuracy increasing over the epochs

⁶<https://colab.research.google.com/drive/1QHIIHhu28sSiWOT-ZGTwg0bp-bulW3011>

3.9. VGG16 ARCHITECTURE

and decline in the loss function. The VGG model was trained using the SGD optimiser and learning rate of 0.01 which resulted in accuracy of 78.14% over the 60 epochs.

3.9.1 Transfer Learning with VGG16

Another, test was performed to evaluate the performance of the model with transferring the weights obtained from the model mentioned above for 300 epochs.

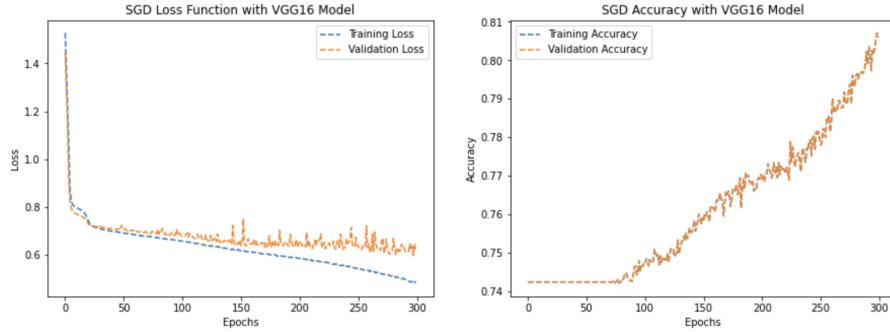


Figure 3.17: VGG16 Model Architecture

In the figure 3.17 the training and validation accuracy was increasing with same growth with increase in the number of epochs. The result obtained from the experiment was 77.48% on testing data which is worst then the evaluation of the previous model. The poor model accuracy could be a result of overfitting of the convolutional network.

3.10 Lenet Architecture

The lenet model was implemented in the keras which requires the input image of $32 * 32$ pixels. As, a result all the pre processing and normalisation was performed on images of pigmented skin lesion on required sizes. The lenet architecture was easy to train and required less time to learn the relationship in the given data because of the light network architecture and images of low dimensions.

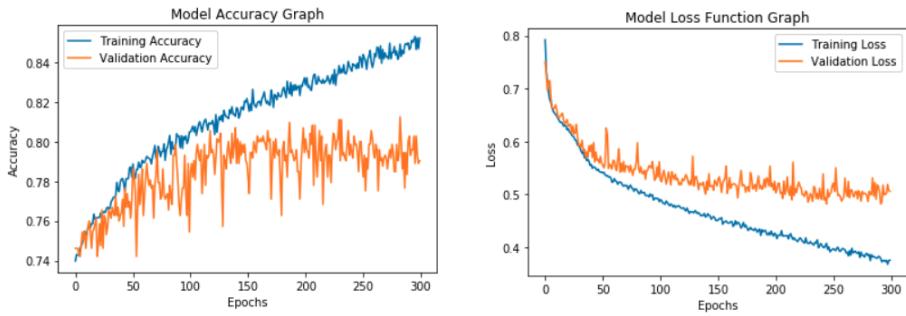


Figure 3.18: Lenet Model training

The figure 3.18 shows the model accuracy and loss function graph of the first experiment performed on the lenet model architecture with SGD optimiser and learning rate of 0.01 for 300 epochs. The result obtained from accuracy of the evaluation performed on the testing data was 80.35%. However, the traditional method of training the lenet model architecture requires using Adam optimiser with learning rate of 0.0001 and resulted in better accuracy evaluation of 81.34% on testing data.

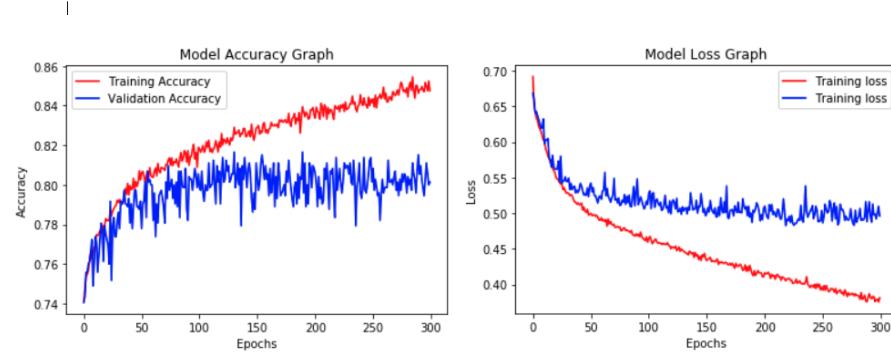
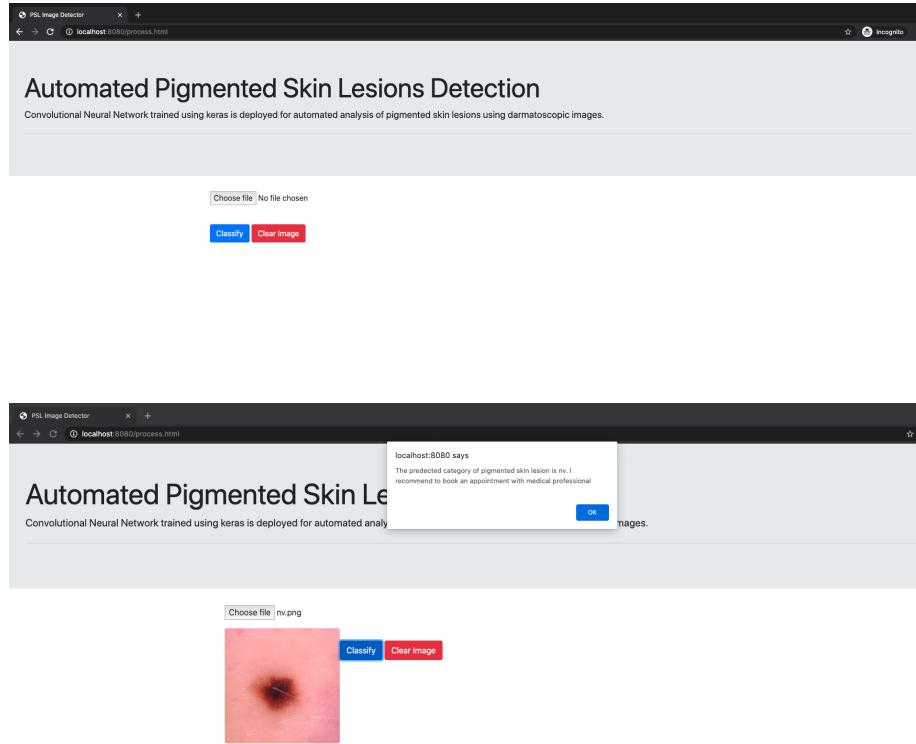


Figure 3.19: Transfer Learning Lenet Model training

3.10. LENET ARCHITECTURE

The figure 3.19 show the model training training performed using the transfer learning on the lenet in attempt to achieve better model accuracy results. The model accuracy obtained from the transfer learning approach on lenet architecture was 81.56% on testing data. Therefore, it was most successful model so, far in analysing the categories of pigmented skin lesions. However, the lenet architecture only functions on small image size of 32 * 32 pixels which when resized might loose vital information in the process of diagnosing pigmented skin lesions.

3.11 Developing Web System



The model with the best accuracy trained in keras was converted and loaded into the tensorflowjs web framework. Furthermore, the website was developed using HTML5, CSS3, Bootstrap and Javascript to get the file uploaded by the client and interact with the model deployed on the node js server which also serves the static contents of the website. The website displays the user with the classification of pigmented skin lesions with most percentage among all the other categories of pigmented skin lesions. The web system was developed as the proof of concept and can further be developed to book appointments with medical professionals.

⁷

⁷<https://github.coventry.ac.uk/sareenv/Final-Year-Project/tree/master/webSystem>

Chapter 4

Evaluations and Results

The analysis was performed based on the data collected from the participants from the medical background. The evaluation section focuses on test the time efficiency of the artificial neural network in predicting pigmented skin lesions. Furthermore, the confusion matrix and f1 scores are used to analyse the performance of the classification.

4.1 Stastical Hypothesis Testing

The following independent unpaired stastical testing was peformed using IBM SPSS software. The H_0 (null hypothesis) = there is no difference in predicting the class of pigmented skin lesions by automated system and medical professional. On the otherhand H_1 (alternative hypothesis) = there is a difference in time required to predict pigmented skin lesions.

T-Test

Group Statistics								
	Group	N	Mean	Std. Deviation	Std. Error Mean			
Time	1	10	.0321	.00604	.00191			
	2	10	246.0000	91.43304	28.91366			
Independent Samples Test								
	Levene's Test for Equality of Variances			t-test for Equality of Means			95% Confidence Interval of the Difference	
	F	Sig.		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Time	Equal variances assumed		14.796	.001	-8.507	18	.000	-245.96793
	Equal variances not assumed				-8.507	9.000	.000	28.91366
							28.91366	-306.71329
							-311.37519	-180.56068

Figure 4.1: Independent sample t-test

4.2. CONFUSION MATRIX

The figure 4.1 shows the results obtained by comparing the means of the time required by the automated system and medical professional to predict pigmented skin lesions. The group 1 in the test is referred to the automated system and group 2 is referred to medical professionals. Based on the p value obtained from the test which is 0.000 as in 4.1 in the Sig(2-tailed) the null hypothesis is failed. Thus, it can be concluded that there is a significant difference in time required to perform diagnosis. The purposed automated machine is more time efficient to predict.

4.2 Confusion Matrix

		Truth		
		Condition positive	Condition negative	
Prediction	Predicted condition positive	True positive	False positive	Positive predictive value = True positive/predicted condition positive
	Predicted condition negative	False negative	True negative	Negative predictive value = True negative/predicted condition negative
		Sensitivity = True positive/condition positive	Specificity = True negative/condition negative	

Confusion Matrix also known as the error matrix is the table to describe the performance of automated system given true values of classification classes (Sharma, 2017). The confusion matrix also allows the misclassified predicted

4.2. CONFUSION MATRIX

labels by the automated and helps to evaluate the performance measures (Sharma, 2017).

Recall and Precision of Automated System

$$\text{Recall} = \frac{\text{TruePositive}}{\text{FalseNegative} + \text{TruePositive}}$$

$$\text{Precision} = \frac{\text{TruePositive}}{\text{TruePositive} + \text{FalsePositive}}$$

The equations above are be used to obtain the precision and recall values of the convolutional neural network. The Precision and recall values are further used to compute the f-measures also known as the f1score to evaluate the performance of automated system.

$$\text{F-measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}}$$

Confusion Matrix and F-measure of the Neural network

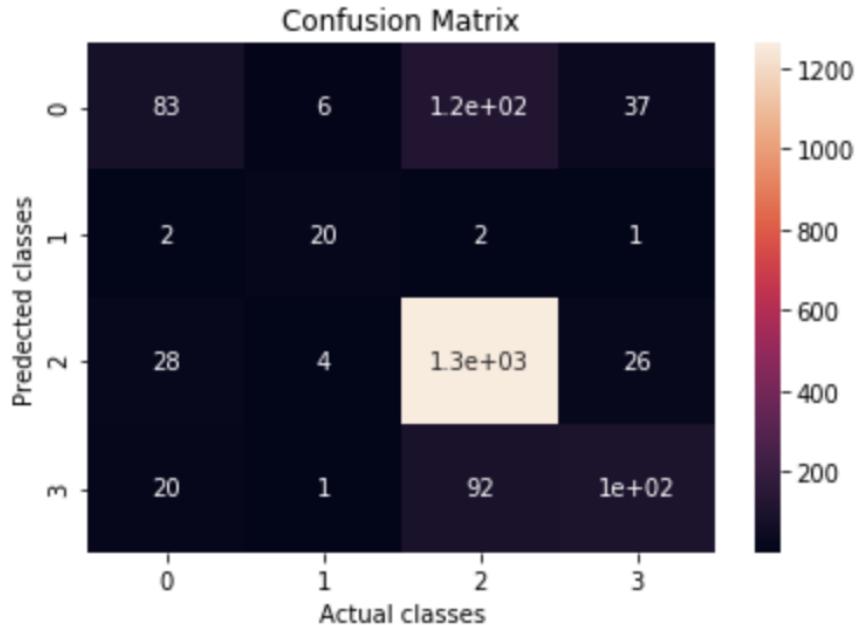


Figure 4.2: Error Matrix

The confusion matrix was computed using the sklearn learn library which requires the predicted values and actual true labels for each pigmented skin lesions. Futhermore the confusion matrix was plotted using heatmaps in the matplotlib visualisation library.

4.2. CONFUSION MATRIX

	precision	recall	f1-score	support
1	0.62	0.33	0.43	251
2	0.65	0.80	0.71	25
3	0.85	0.96	0.90	1319
4	0.62	0.48	0.54	217
accuracy			0.81	1812
macro avg	0.69	0.64	0.65	1812
weighted avg	0.79	0.81	0.79	1812

Figure 4.3: Classification Report of automated system

The figure above shows the analysis of the automated system in an attempt to predict pigmented skin lesions. The report above shows precision, recall and f1 score values which indicates the performance in predicting each category of pigmented skin lesions. All further analysis to compare the model performance with participants are evaluated based on the f1 scores for each category of pigmented skin lesion.

4.3. MOST COMMON METHOD FOR DIAGNOSING SKIN LESIONS

4.3 Most common method for diagnosing skin lesions

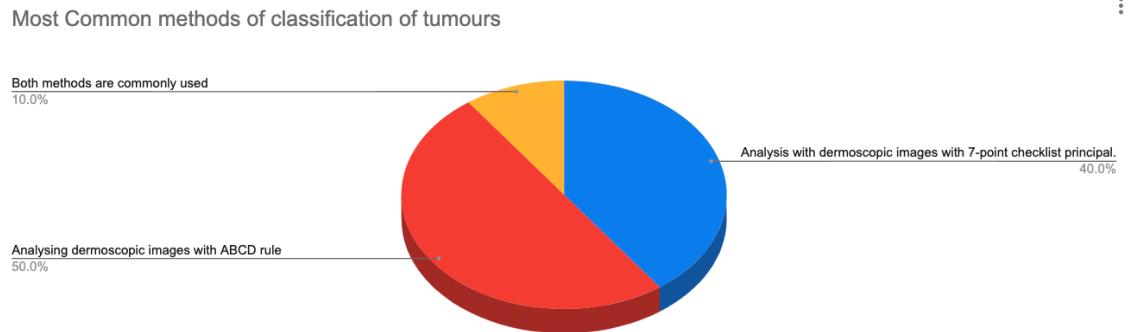
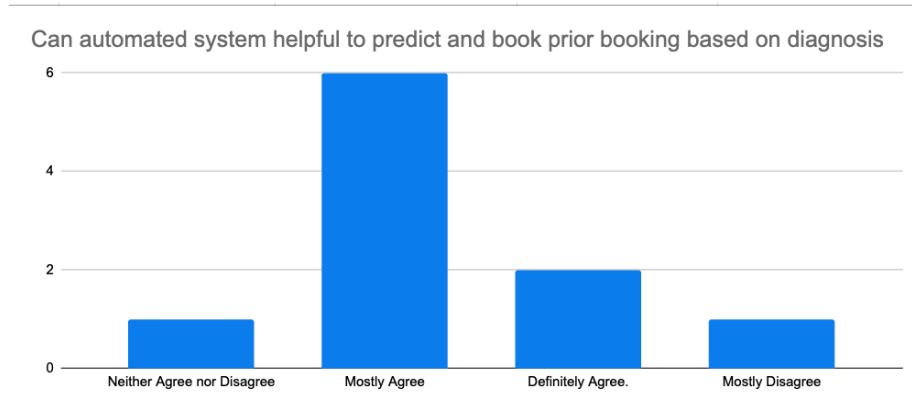


Figure 4.4: Most Common Diagnosis

Figure 4.4 shows the results obtained from the participants on most common method of diagnosing skin lesions. The half majority of the overall participants confirmed that ABCD rule is used to detect pigmented skin lesions. Whereas 40% and 10% population believed in diagnosing using 7 checklist principal method and combination of both the technique respectively.

4.4 Automated System can help to patients to diagnose book prior appointments



Based on the results obtained from participants as shown in the bar chart above majority of the participants mostly agree that automated system can help

4.4. AUTOMATED SYSTEM CAN HELP TO PATIENTS TO DIAGNOSE BOOK PRIOR APPOINTMENTS

patients to book prior appointments and to predict classes of pigmented skin lesions. Whereas, two participants were mostly disagree and Neither Agree nor Disagree to consider the system helpful for general audience to operate.

4.5 Classification comparison with Participants

The following evaluations are performed to compare the classification performance of participants and automated system and generate analysis report.

Classification comparison with Participant 1

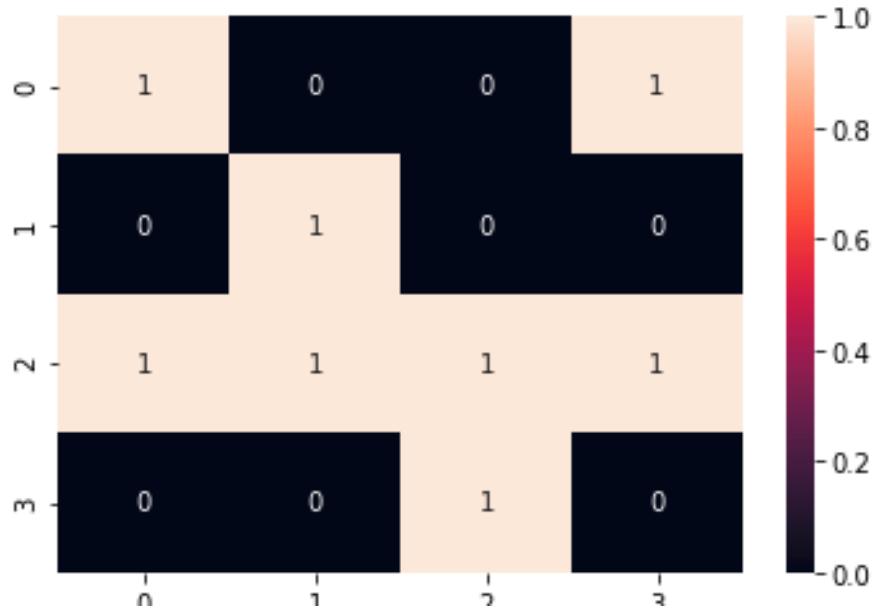


Figure 4.5: Confusion Matrix from Participant 1

	precision	recall	f1-score	support
1	0.50	0.50	0.50	2
2	0.50	1.00	0.67	1
3	0.50	0.25	0.33	4
4	0.00	0.00	0.00	1
accuracy			0.38	8
macro avg	0.38	0.44	0.37	8
weighted avg	0.44	0.38	0.38	8

Figure 4.6: Classification Report of participant 1

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

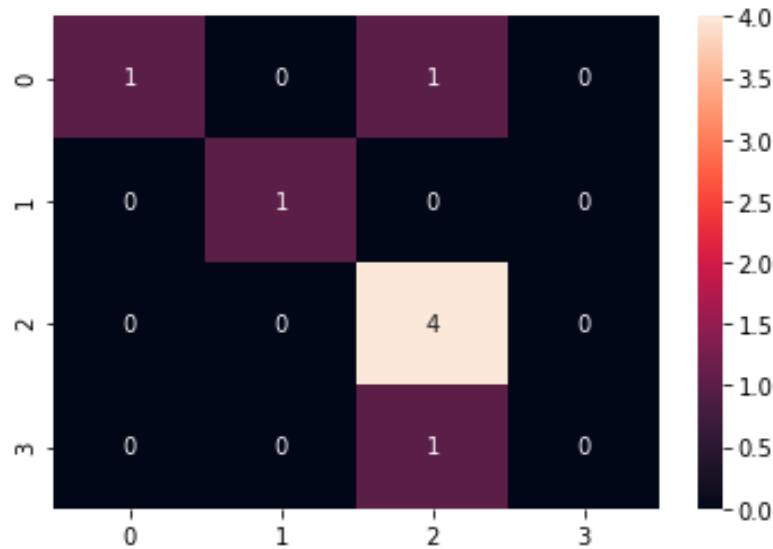


Figure 4.7: Confusion Matrix from Automated System prediction

	precision	recall	f1-score	support
1	1.00	0.50	0.67	2
2	1.00	1.00	1.00	1
3	0.67	1.00	0.80	4
4	0.00	0.00	0.00	1
accuracy			0.75	8
macro avg	0.67	0.62	0.62	8
weighted avg	0.71	0.75	0.69	8

Figure 4.8: Classification Report of Automated System

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

Classification comparison with Participant 2

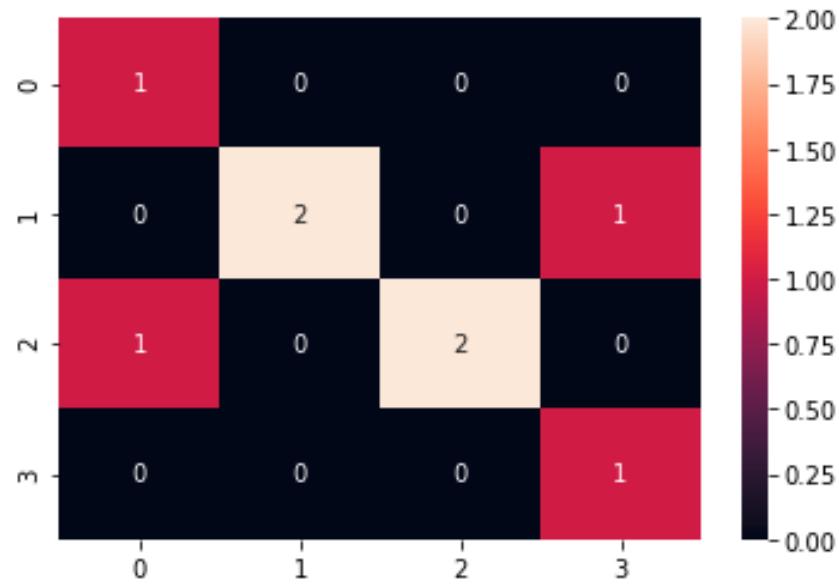


Figure 4.9: Confusion Matrix from Participant 2

	precision	recall	f1-score	support
1	0.50	1.00	0.67	1
2	1.00	0.67	0.80	3
3	1.00	0.67	0.80	3
4	0.50	1.00	0.67	1
accuracy			0.75	8
macro avg	0.75	0.83	0.73	8
weighted avg	0.88	0.75	0.77	8

Figure 4.10: Classification Report of participant 2

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

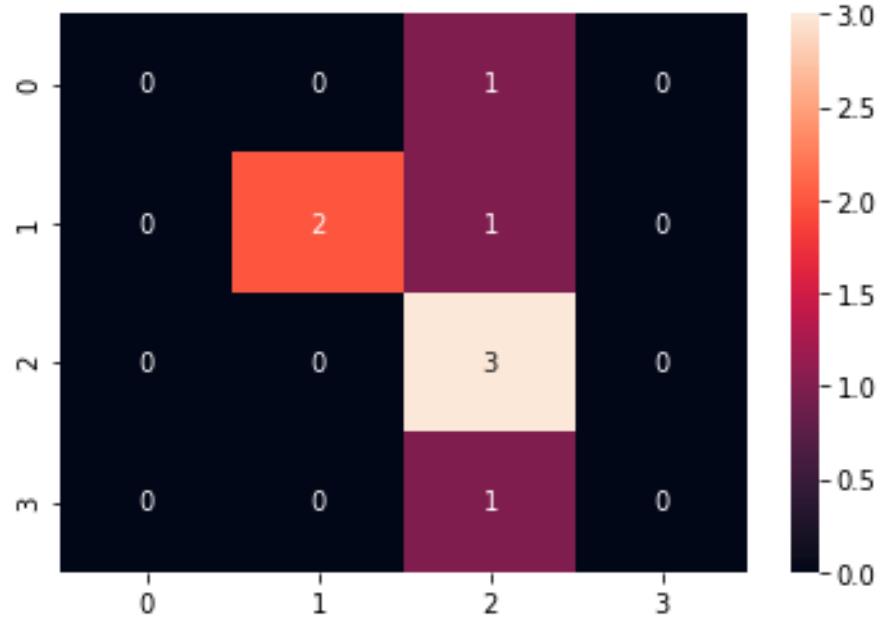


Figure 4.11: Confusion Matrix from Automated System prediction

	precision	recall	f1-score	support
1	0.00	0.00	0.00	1
2	1.00	0.67	0.80	3
3	0.50	1.00	0.67	3
4	0.00	0.00	0.00	1
accuracy			0.62	8
macro avg	0.38	0.42	0.37	8
weighted avg	0.56	0.62	0.55	8

Figure 4.12: Classification Report of Automated System

Classification comparison with Participant 3

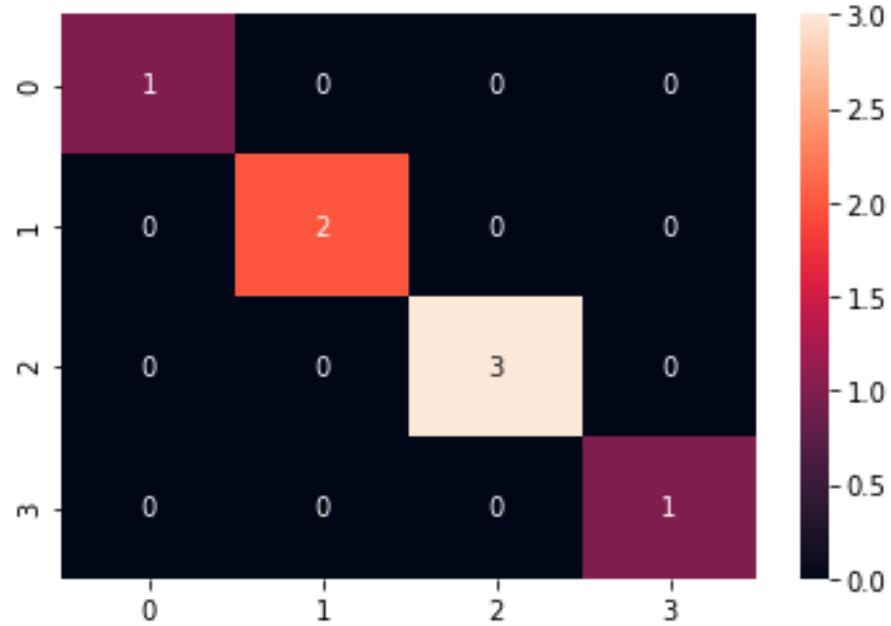


Figure 4.13: Confusion Matrix from Participant 3

	precision	recall	f1-score	support
1	1.00	1.00	1.00	1
2	1.00	1.00	1.00	2
3	1.00	1.00	1.00	3
4	1.00	1.00	1.00	1
accuracy			1.00	7
macro avg	1.00	1.00	1.00	7
weighted avg	1.00	1.00	1.00	7

Figure 4.14: Classification Report of participant 3

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

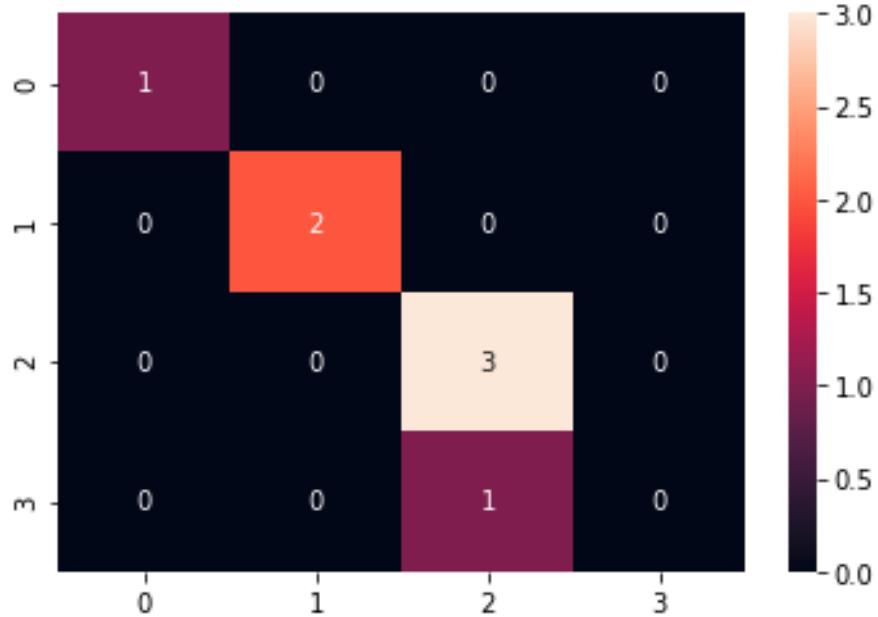


Figure 4.15: Confusion Matrix from Automated System prediction

	precision	recall	f1-score	support
1	1.00	1.00	1.00	1
2	1.00	1.00	1.00	2
3	0.75	1.00	0.86	3
4	0.00	0.00	0.00	1
accuracy			0.86	7
macro avg	0.69	0.75	0.71	7
weighted avg	0.75	0.86	0.80	7

Figure 4.16: Classification Report of Automated System

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

Classification comparison with Participant 4

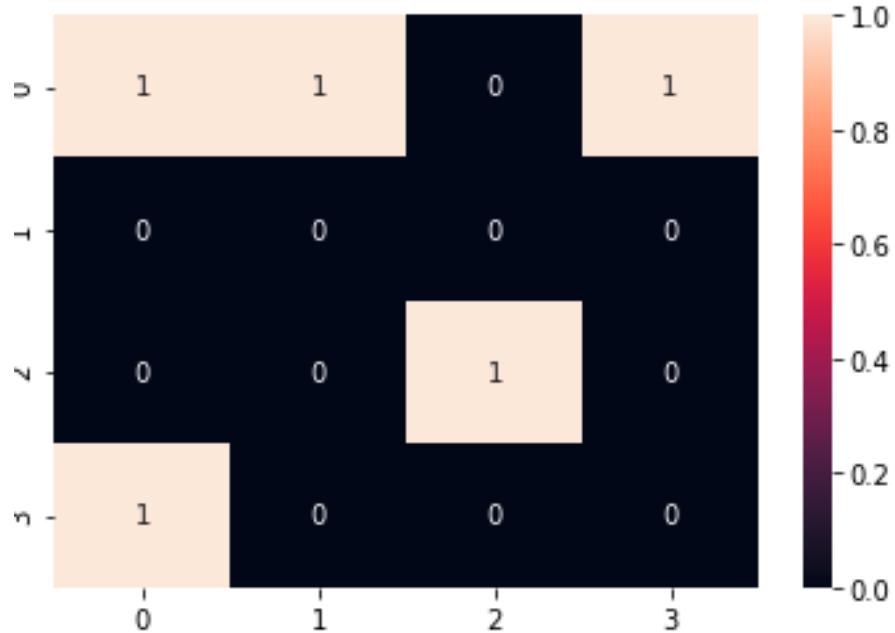


Figure 4.17: Confusion Matrix from Participant 4

	precision	recall	f1-score	support
1	0.50	0.33	0.40	3
2	0.00	0.00	0.00	0
3	1.00	1.00	1.00	1
4	0.00	0.00	0.00	1
accuracy			0.40	5
macro avg	0.38	0.33	0.35	5
weighted avg	0.50	0.40	0.44	5

Figure 4.18: Classification Report of participant 4

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

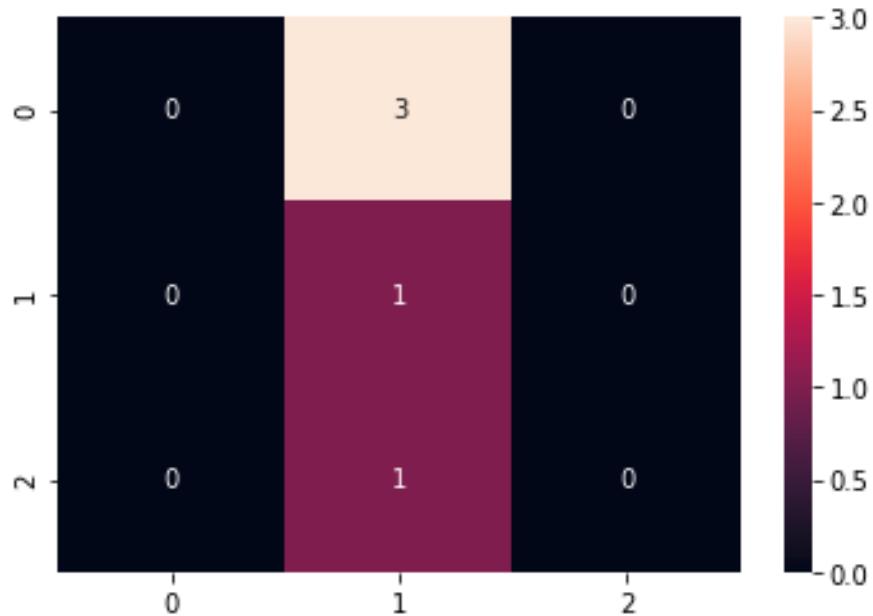


Figure 4.19: Confusion Matrix from Automated System prediction

	precision	recall	f1-score	support
1	0.00	0.00	0.00	3
3	0.20	1.00	0.33	1
4	0.00	0.00	0.00	1
accuracy			0.20	5
macro avg	0.07	0.33	0.11	5
weighted avg	0.04	0.20	0.07	5

Figure 4.20: Classification Report of Automated System

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

Classification comparison with Participant 5

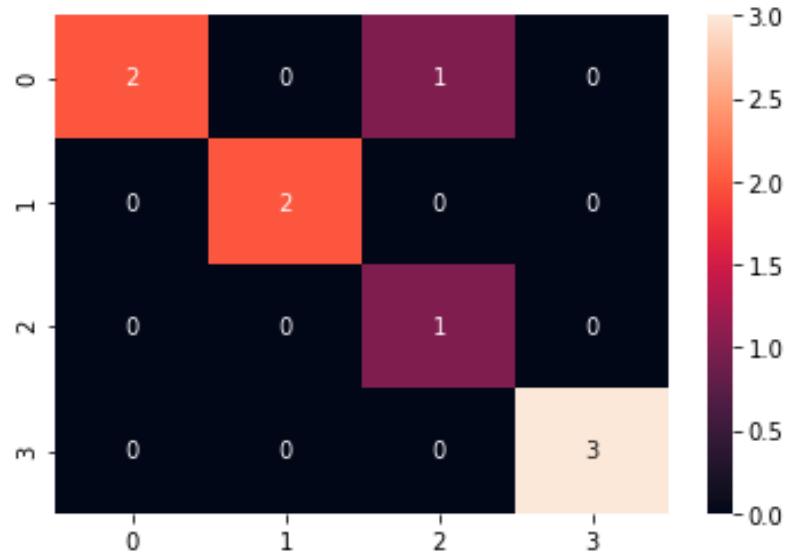


Figure 4.21: Confusion Matrix from Participant 5

	precision	recall	f1-score	support
1	1.00	0.67	0.80	3
2	1.00	1.00	1.00	2
3	0.50	1.00	0.67	1
4	1.00	1.00	1.00	3
accuracy			0.89	9
macro avg	0.88	0.92	0.87	9
weighted avg	0.94	0.89	0.90	9

Figure 4.22: Classification Report of participant 5

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

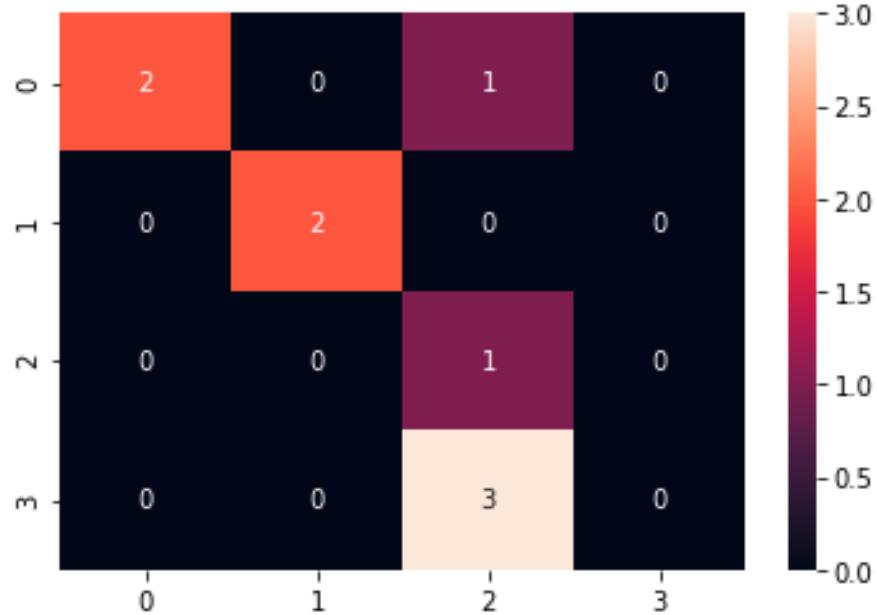


Figure 4.23: Confusion Matrix from Automated System prediction

	precision	recall	f1-score	support
1	1.00	0.67	0.80	3
2	1.00	1.00	1.00	2
3	0.20	1.00	0.33	1
4	0.00	0.00	0.00	3
accuracy			0.56	9
macro avg	0.55	0.67	0.53	9
weighted avg	0.58	0.56	0.53	9

Figure 4.24: Classification Report of Automated System

Classification comparison with Participant 6

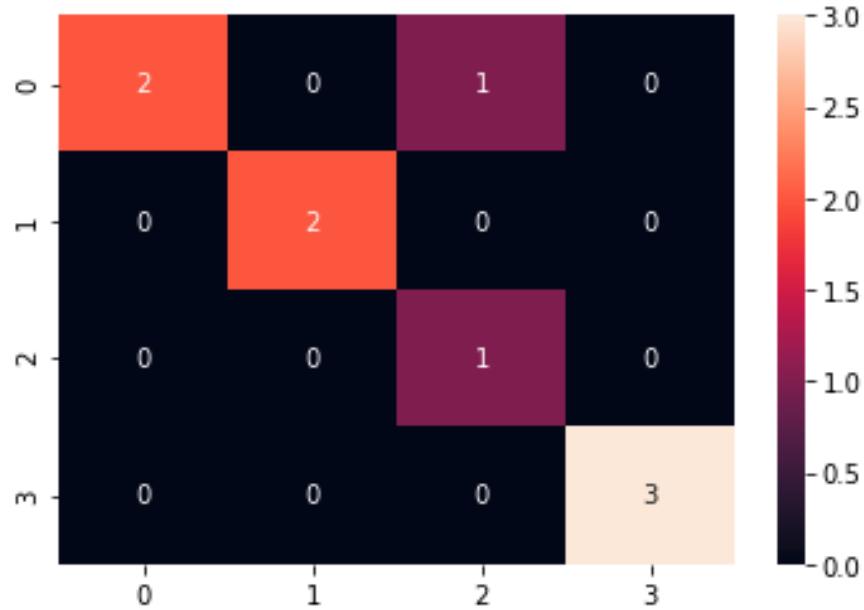


Figure 4.25: Confusion Matrix from Participant 6

	precision	recall	f1-score	support
1	1.00	0.67	0.80	3
2	1.00	1.00	1.00	2
3	0.50	1.00	0.67	1
4	1.00	1.00	1.00	3
accuracy			0.89	9
macro avg	0.88	0.92	0.87	9
weighted avg	0.94	0.89	0.90	9

Figure 4.26: Classification Report of participant 6

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

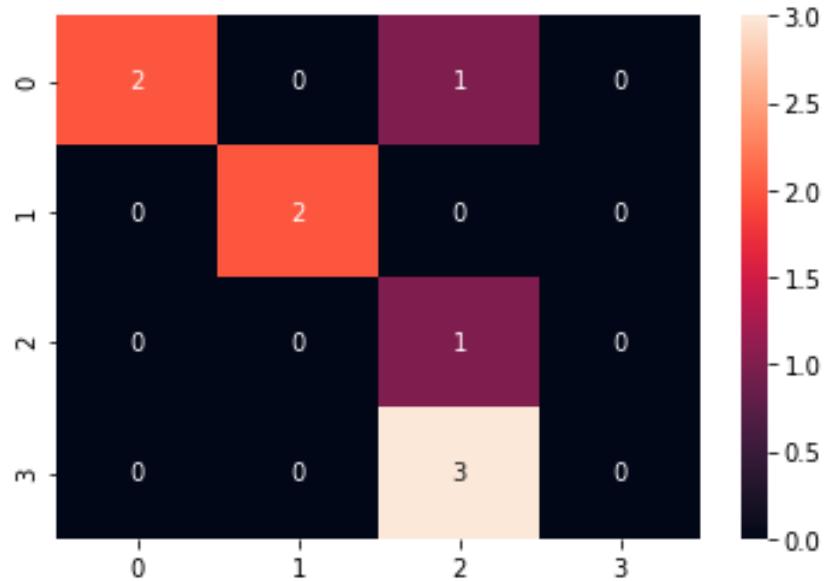


Figure 4.27: Confusion Matrix from Automated System prediction

	precision	recall	f1-score	support
1	1.00	0.67	0.80	3
2	1.00	1.00	1.00	2
3	0.20	1.00	0.33	1
4	0.00	0.00	0.00	3
accuracy			0.56	9
macro avg	0.55	0.67	0.53	9
weighted avg	0.58	0.56	0.53	9

Figure 4.28: Classification Report of Automated System

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

Classification comparison with Participant 7

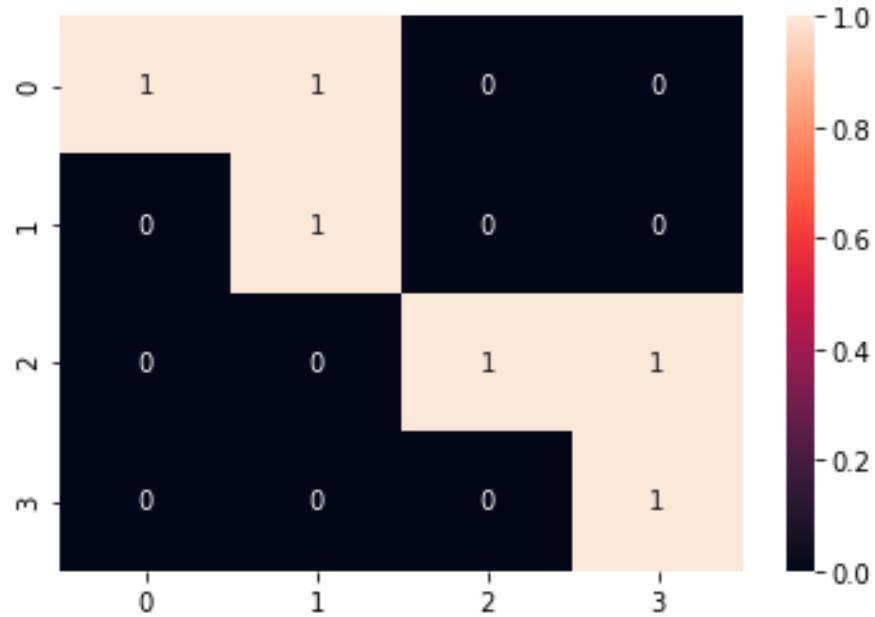


Figure 4.29: Confusion Matrix from Participant 7

	precision	recall	f1-score	support
1	1.00	0.50	0.67	2
2	0.50	1.00	0.67	1
3	1.00	0.50	0.67	2
4	0.50	1.00	0.67	1
accuracy			0.67	6
macro avg	0.75	0.75	0.67	6
weighted avg	0.83	0.67	0.67	6

Figure 4.30: Classification Report of participant 7

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

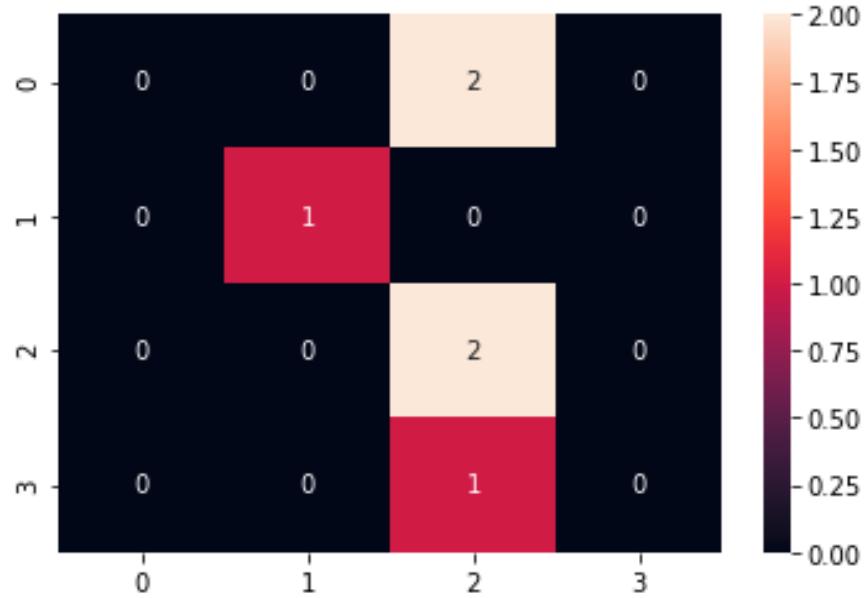


Figure 4.31: Confusion Matrix from Automated System prediction

	precision	recall	f1-score	support
1	0.00	0.00	0.00	2
2	1.00	1.00	1.00	1
3	0.40	1.00	0.57	2
4	0.00	0.00	0.00	1
accuracy			0.50	6
macro avg	0.35	0.50	0.39	6
weighted avg	0.30	0.50	0.36	6

Figure 4.32: Classification Report of Automated System

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

Classification comparison with Participant 8

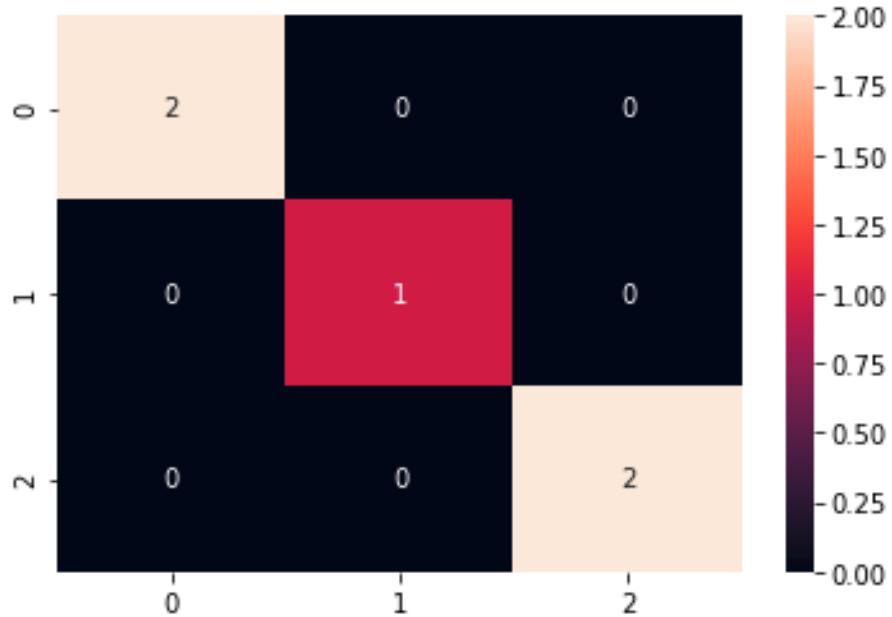


Figure 4.33: Confusion Matrix from Participant 8

	precision	recall	f1-score	support
2	1.00	1.00	1.00	2
3	1.00	1.00	1.00	1
4	1.00	1.00	1.00	2
accuracy			1.00	5
macro avg	1.00	1.00	1.00	5
weighted avg	1.00	1.00	1.00	5

Figure 4.34: Classification Report of participant 8

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

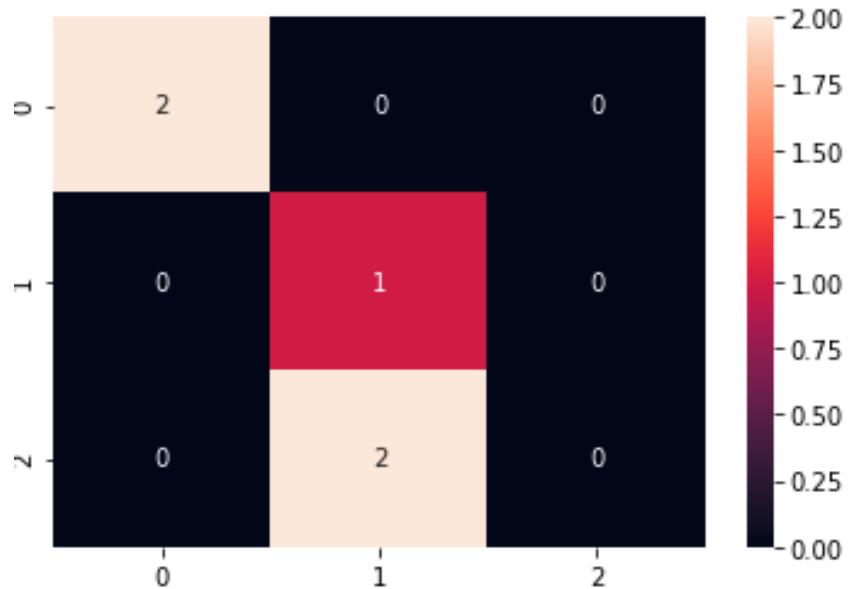


Figure 4.35: Confusion Matrix from Automated System prediction

	precision	recall	f1-score	support
2	1.00	1.00	1.00	2
3	0.33	1.00	0.50	1
4	0.00	0.00	0.00	2
accuracy			0.60	5
macro avg	0.44	0.67	0.50	5
weighted avg	0.47	0.60	0.50	5

Figure 4.36: Classification Report of Automated System

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

Classification comparison with Participant 9

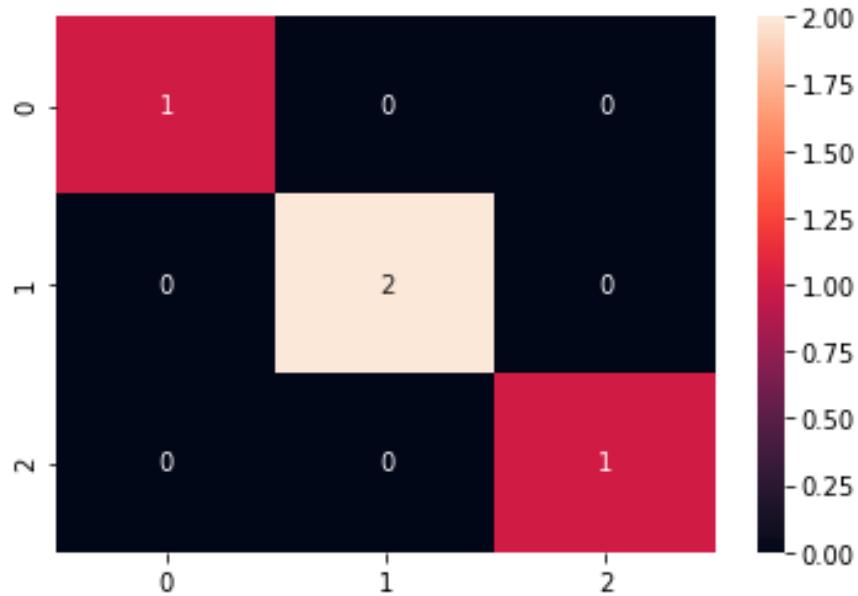


Figure 4.37: Confusion Matrix from Participant 9

	precision	recall	f1-score	support
1	1.00	1.00	1.00	1
2	1.00	1.00	1.00	2
4	1.00	1.00	1.00	1
accuracy			1.00	4
macro avg	1.00	1.00	1.00	4
weighted avg	1.00	1.00	1.00	4

Figure 4.38: Classification Report of participant 9

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

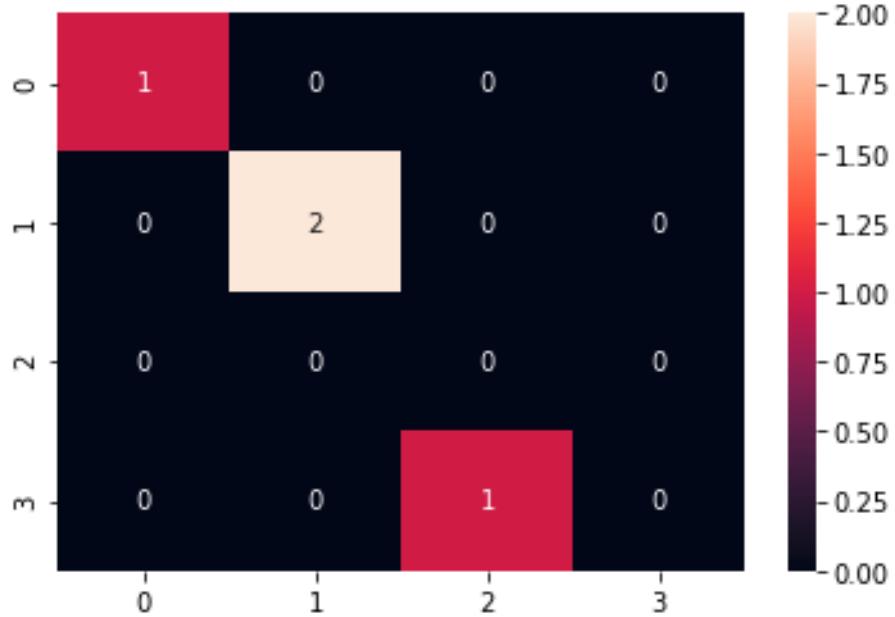


Figure 4.39: Confusion Matrix from Automated System prediction

	precision	recall	f1-score	support
1	1.00	1.00	1.00	1
2	1.00	1.00	1.00	2
3	0.00	0.00	0.00	0
4	0.00	0.00	0.00	1
accuracy			0.75	4
macro avg	0.50	0.50	0.50	4
weighted avg	0.75	0.75	0.75	4

Figure 4.40: Classification Report of Automated System

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

Classification comparison with Participant 10

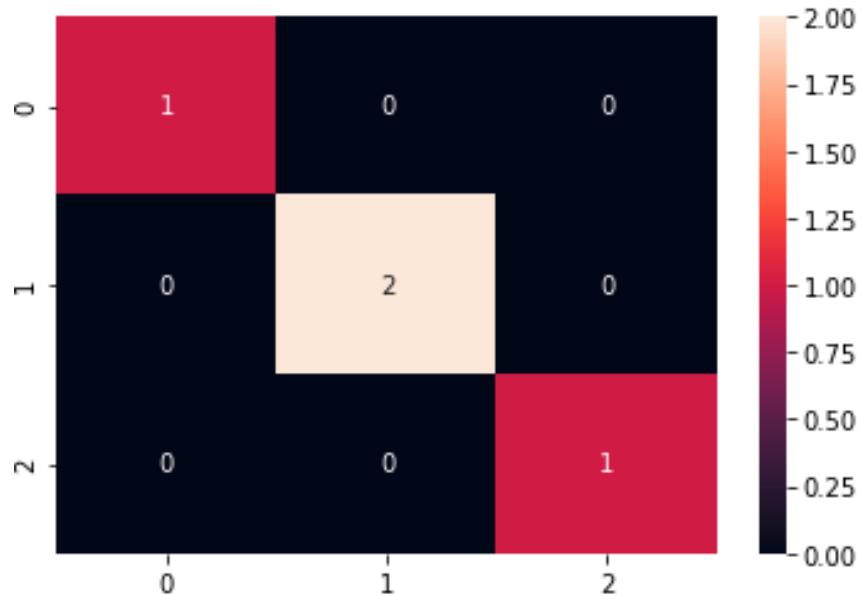


Figure 4.41: Confusion Matrix from Participant 10

	precision	recall	f1-score	support
1	1.00	1.00	1.00	1
2	1.00	1.00	1.00	2
4	1.00	1.00	1.00	1
accuracy			1.00	4
macro avg	1.00	1.00	1.00	4
weighted avg	1.00	1.00	1.00	4

Figure 4.42: Classification Report of participant 10

4.5. CLASSIFICATION COMPARISON WITH PARTICIPANTS

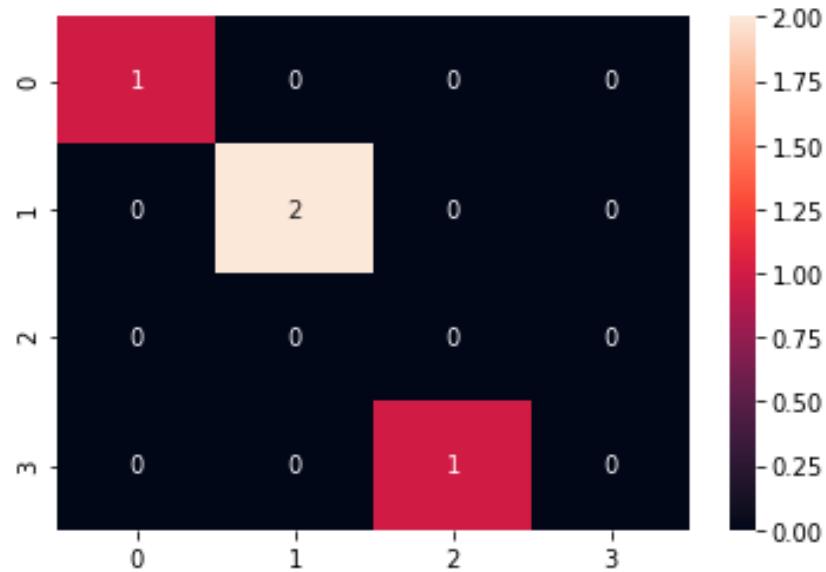


Figure 4.43: Confusion Matrix from Automated System prediction

	precision	recall	f1-score	support
1	1.00	1.00	1.00	1
2	1.00	1.00	1.00	2
3	0.00	0.00	0.00	0
4	0.00	0.00	0.00	1
accuracy			0.75	4
macro avg	0.50	0.50	0.50	4
weighted avg	0.75	0.75	0.75	4

Figure 4.44: Classification Report of Automated System

Chapter 5

Discussion

5.1 Achievements of Research Project

The study has succeeded in developing a deep learning-based solution to detect skin cancers from pigmented skin lesions. Different model experiments were performed to analyse the impact of architecture and hyperparameters on the accuracy of the model to identify the problem. The dataset was divided into separate testing data samples which were used for only evaluation purpose. The current research purpose model with the highest accuracy of 81.56% performed on 11,00 data samples of pigmented skin lesions. The comparison was performed on the diagnosis by an automated system and medical professional to understand the time efficiency and reliability of the automated system. The results obtained from performing the comparison has shown that the automated system is significant time-efficient to perform diagnosis. In addition, the model has performed diagnosis to detect skin cancer from dermatoscopic images of pigmented skin lesion better than medical practitioners in some cases. Furthermore, the intelligent model was also deployed on the web-based system to be used by the general audience. The research has helped me understanding the functioning of convolutional neural networks in visual recognition of the pigmented skin lesions and conducting research while considering participants privacy.

5.2 Deficiencies of Research Project

The primary deficiency of the research is the limited collection of the responses for secondary research of the project. The widespread of pandemic coronavirus, has affected data collection from medical institutes ensuring the safety of the author of this research project. However, the small number of datasamples were collected from ten medical professionals to compare model performance. The research experiments were constraint by the available hardware resources of 8 gigabytes of random access memory as result images of lower size and dimensions were used for image preprocessing and normalisation. The research

5.3. FUTURE IMPROVEMENT

currently focuses on identifying the valid pigmented skin lesion out of the limited categories of the classes and be extended with the availability of the data.

5.3 Future Improvement

The results obtained from the research can be used to understand the accurate hyper-parameters for detecting pigmented skin lesions. The current research project can further be improved by analysing the impact of applying the UNET semantic segmentation model which can result in potentially result in better accuracy of the model performance. The web system can be improved by developing the medical appointment booking system based on availability of medical professionals. More data samples need to be collected from medical professionals in the further research for perform the comparison between medical professional and automated system.

Chapter 6

Project Evaluation

6.1 Supervisor Feedback

1. Testing and Evaluation Section

- Confusion matrix also known error matrix was appropriate based on the feed back to compare the results of automated system and medical professionals.
 - The t-tests were recommended by the supervisor for the evaluation of significance of the obtained results.
 - It was advised to get further guidance on conducting the appropriate stastically testing from Sigma maths support center for the evaluation of the results.
-
- **Action Taken :** Independent t-test was peformed for the stastically significant result obtained from evaluation with the support of maths support center using SPSS software.

2. Including model accuracy for each experiments.

- The methodology section was missing the accuracy results obtained from the different model experiments.
-
- **Action Taken :** All the accuracy tables were added to the experiments.

3. Refrancing and Formatting.

- It was advised that relevant links should be included in the footer of the document to provide clear reading experience.
-
- **Action Taken :** All the relevant urls were shifted to the footer using footnote command in latex.

6.2. LEGAL AND ETHICAL ISSUES

6.2 Legal and Ethical Issues

The research project was conducted with ethical considerations where safety and well being of researcher and participants was considered seriously. The Ethics Board has approved the research project for developing a convolutional neural network to detect pigmented skin lesions with ethics id P101878. Due to the current pandemic spread of coronavirus, it was not safe to visit hospitals. As a result, the data collection was performed through the online exchange of forms. The research project involves detecting skin cancer from pigmented skin lesions for which publicly available dataset was used with any personal identity of any patient information which might have ethical considerations to collect data from medical professionals. The participants of the study were completely aware of the purpose of the study. The data collection process of secondary research was compliant with GDPR privacy laws in the united kingdom. The documents containing information collected from participants were stored in the password-protected document and will be destroyed after the research submission. The participants of the research project were informed that they could withdraw their data at any time without any reason or mental pressure. The consent formed was attached to the questionnaire, which required participants to give permissions to collect the data and related concerns.

6.3 Research Challenges

There were various challenges while conducting the research to develop and compare the automated system for classification of pigmented skin lesions. The risk of no prior knowledge about the convolutional neural network was present in the initial phase which required the personal research on foundations of artifical neural network and learning to operate the keras deep learning library. The research involved performing the image processing and normalisation tasks which requires the superior hardware the 8 gigabytes of the RAM (random access memory) was not sufficient to normalise the original images from the dataset and train the VGG16 network. As, a result the images were resized to 224 * 224 pixels to reduce the consumption of the primary memory. Furthermore, the VGG16 model was trained on the google colab platform which provides free computation power to train the artifical neural networks. However, the colab platform requires the user to interact with the web interface to avoid timeout error. The secondary research for the project involves comparing the data collected from the medical professionals with automated system. However, due to the pandemic coronavirus it was not safe to collect the data from medical hospitals which resulted in fewer resulted in fewer responses from the medical professionals.

6.4 Project Management

The dissertation project was managed using the waterfall software development lifecycle and the flexible kanban system to ensure the project timeline is managed properly across various phases of development lifecycle.

6.4.1 Software Development Lifecycle

The waterfall software development lifecycle was used during developing the project. The waterfall model also known as linear-sequential life cycle model focuses on completing each phase of development before switching to the next stage of development (TutorialsPoint, 2015). The waterfall model was accurate

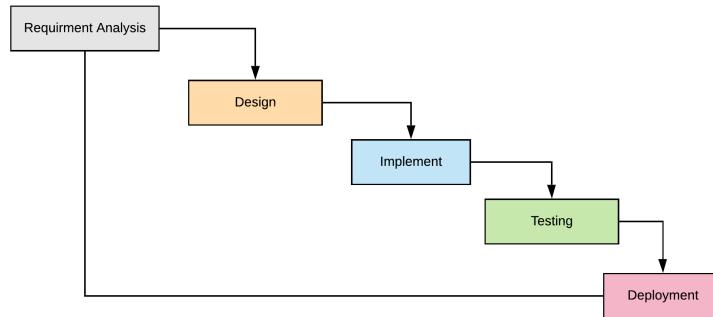


Figure 6.1: Waterfall Model

for this project as the scope of the project is small and was developed individually. Output from each phase of development lifecycle acts as the input to the next phase. The figure above 6.1 shows the various phases of the waterfall development lifecycle. In initial phase of the development the requirements and targeted audience were analysed as mentioned in the introduction section of the project. The second phase involved developing conceptual models using UML(Unified Modelling Language) to design the system using component diagram which captures the structural and sequence diagram to capture the behavioural design of the system. The next phase of waterfall model was implementation of the intelligent system to detect pigmented skin lesions. The implementation phase of the lifecycle involved developing convolutional network and performing various experiments to get optimal performance. Furthermore, The next phase of development was to test the effectiveness of the overall system for which the prediction results gathered from medical professionals in the form of questionnaires were compared with the automated system and results were analysed with confusion matrix. The last phase was to deploy the model to the web-based system which was done by converting the keras model to json file and using tensorflowjs. At last, the maintenance phase of lifecycle cannot be applied to short term project.

6.4. PROJECT MANAGEMENT

6.4.2 Kanban Work Management System

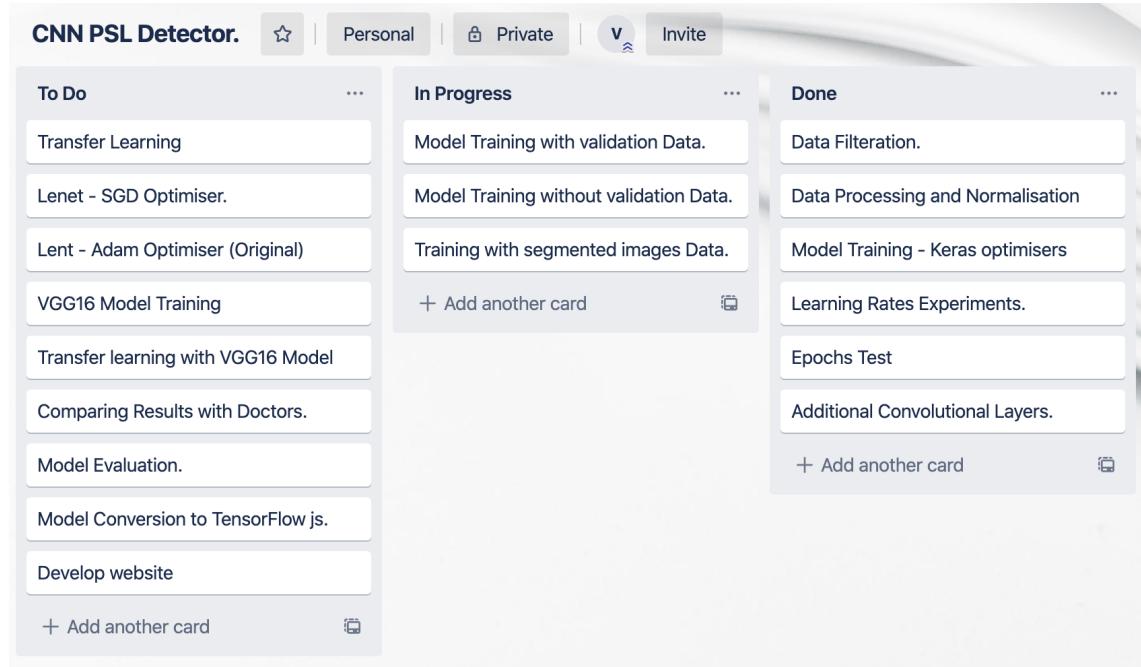


Figure 6.2: Kanban Board

Kanban board was used as the work management system to track the visual progress of the project as shown in the figure 6.2. Kanban is a visual system to manage the workflow and identifying the process bottlenecks to fix in time efficient manner(Digite, 2017). Kanban system was first introduced in 1940s by Taiichi for toyota automotive in japan (Digite, 2017). The system provides the flexibility while working on project where timeline is difficult to anticipate. Trello platform was used during the research project for kanban system approach to manage the processes and access the visual interface for outstanding processes.

6.4.3 Version Control System

Git version control system was used during the development of different experiments and web based system to ensure the backup of the work done. All the implementation of convolutional neural networks and other related work is available on the master branch of the github coventry university repository. The project does not use any git workflow to maintain the simplicity as the research was conducted individually.¹

¹<https://github.coventry.ac.uk/sareenv/Final-Year-Project>

Chapter 7

Conclusion

The research was successful in achieving the primary objective to develop a deep learning-based solution to detect skin cancer. The research has compared the diagnosis by the medical professionals and automated system. Furthermore, the automated deep learning-based solution can significantly be time-efficient and help in the medical sector to decrease mortality concerning skin cancer.

Bibliography

- Agatonovic-Kustrin, S. and Beresford, R. (2000), ‘Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research’, *Journal of Pharmaceutical and Biomedical Analysis* **22**(5), 717 – 727.
URL: <http://www.sciencedirect.com/science/article/pii/S0731708599002721>
- Al-Amri, S. S., Kalyankar, N. V. et al. (2010), ‘Image segmentation by using threshold techniques’, *arXiv preprint arXiv:1005.4020* .
- Bono, A., Bartoli, C., Baldi, M., Moglia, D., Tomatis, S., Tragni, G., Cascinelli, N. and Santinami, M. (2004), ‘Micro-melanoma detection. a clinical study on 22 cases of melanoma with a diameter equal to or less than 3 mm’, *Tumori Journal* **90**(1), 128–131. PMID: 15143985.
URL: <https://doi.org/10.1177/030089160409000125>
- Celebi, M. E., Kingravi, H. A., Uddin, B., Iyatomi, H., Aslandogan, Y. A., Stoecker, W. V. and Moss, R. H. (2007), ‘A methodological approach to the classification of dermoscopy images’, *Computerized Medical Imaging and Graphics* **31**(6), 362 – 373.
URL: <http://www.sciencedirect.com/science/article/pii/S0895611107000146>
- Daniel Jensen, J. and Elewski, B. E. (2015), ‘The abcdef rule: Combining the ”abcde rule” and the ”ugly duckling sign” in an effort to improve patient self-screening examinations’, *The Journal of clinical and aesthetic dermatology* **8**(2), 15–15. 25741397[pmid].
URL: <https://pubmed.ncbi.nlm.nih.gov/25741397>
- Digite (2017), Online. Accessed 03-2020.
URL: <https://www.digite.com/kanban/>
- Garnavi, R., Aldeen, M., Celebi, M. E., Varigos, G. and Finch, S. (2011), ‘Border detection in dermoscopy images using hybrid thresholding on optimized color channels’, *Computerized Medical Imaging and Graphics* **35**(2), 105 – 115. Advances in Skin Cancer Image Analysis.
URL: <http://www.sciencedirect.com/science/article/pii/S0895611110000819>

BIBLIOGRAPHY

- Grob, J. and Bonerandi, J. (1998), ‘The ‘ugly duckling’ sign: identification of the common characteristics of nevi in an individual as a basis for melanoma screening’, *Archives of dermatology* **134**(1), 103–104.
- Kawahara, J., BenTaieb, A. and Hamarneh, G. (2016), Deep features to classify skin lesions, in ‘2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)’, pp. 1397–1400.
- Korotkov, K. and Garcia, R. (2012), ‘Computerized analysis of pigmented skin lesions: A review’, *Artificial Intelligence in Medicine* **56**(2), 69 – 90.
URL: <http://www.sciencedirect.com/science/article/pii/S0933365712001108>
- Kussul, E., Baidyk, T., Kasatkina, L. and Lukovich, V. (2001), Rosenblatt perceptrons for handwritten digit recognition, in ‘IJCNN’01. International Joint Conference on Neural Networks. Proceedings (Cat. No.01CH37222)’, Vol. 2, pp. 1516–1520 vol.2.
- LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998), ‘Gradient-based learning applied to document recognition’, *Proceedings of the IEEE* **86**(11), 2278–2324.
- Liew, S. S., Hani, M. K., Radzi, S. A. and Bakhteri, R. (2016), ‘Gender classification: a convolutional neural network approach’, *Turkish Journal of Electrical Engineering & Computer Sciences* **24**(3), 1248–1264.
- Loescher, L. J., Janda, M., Soyer, H. P., Shea, K. and Curiel-Lewandrowski, C. (2013), ‘Advances in skin cancer early detection and diagnosis’, *Seminars in Oncology Nursing* **29**(3), 170 – 181. Skin Cancer.
URL: <http://www.sciencedirect.com/science/article/pii/S0749208113000338>
- Lubis, F. F., Rosmansyah, Y. and Supangkat, S. H. (2014), Gradient descent and normal equations on cost function minimization for online predictive using linear regression with multiple variables, in ‘2014 International Conference on ICT For Smart Society (ICISS)’, pp. 202–205.
- O’Shea, K. and Nash, R. (2015), ‘An introduction to convolutional neural networks’, *CoRR abs/1511.08458*.
URL: <http://arxiv.org/abs/1511.08458>
- Patterson, J. and Gibson, A. (2017), *Deep learning: A practitioner’s approach*, ” O’Reilly Media, Inc.”.
- Sharma, A. (2017), ‘Confusion matrix in machine learning’, Online. Accessed 03-2020.
URL: <https://www.digite.com/kanban/>
- Thompson, F. and Jeyakumar, M. (2017), ‘Vector based classification of dermoscopic images using surf’, *International Journal of Applied Engineering Research* **12**(8), 1758–1764.

BIBLIOGRAPHY

Tschandl, P. (2018), ‘The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions’.

URL: <https://doi.org/10.7910/DVN/DBW86T>

TutorialsPoint (2015), Online. Accessed 03-2020.

URL: <https://www.tutorialspoint.com/sdlc>

Appendices

Participant No.

INFORMED CONSENT FORM:

Convolutional Neural Network-based medical check-up system for Pigmented Skin Lesions Classification.

You are invited to take part in this research study for the purpose of collecting data on evaluating the reliability of automated skin check-up system for classification of common types of pigmented skin lesions.

Before you decide to take part, you must read the accompanying Participant Information Sheet.

Please do not hesitate to ask questions if anything is unclear or if you would like more information about any aspect of this research. It is important that you feel able to take the necessary time to decide whether or not you wish to take part.

If you are happy to participate, please confirm your consent by circling YES against each of the below statements and then signing and dating the form as participant.

1	I confirm that I have read and understood the <u>Participant Information Sheet</u> for the above study and have had the opportunity to ask questions	YES	NO
2	I understand my participation is voluntary and that I am free to withdraw my data, without giving a reason, by contacting the lead researcher and the Research Support Office <u>at any time</u> until the date specified in the Participant Information Sheet	YES	NO
3	I have noted down my participant number (top left of this Consent Form) which may be required by the lead researcher if I wish to withdraw from the study	YES	NO
4	I understand that all the information I provide will be held securely and treated confidentially	YES	NO
5	I am happy for the information I provide to be used (anonymously) in academic papers and other formal research outputs	YES	NO
6	I am happy to answer questions asked in this <u>questionnaire</u>	YES	NO
7	I agree to take part in the above study	YES	NO

Thank you for your participation in this study. Your help is very much appreciated.

Participant's Name	Date	Signature
Researcher	Date	Signature

Consent form



Participation Information Sheet

Convolutional Neural Network-based medical check-up system for Pigmented Skin Lesions Classification.

PARTICIPANT INFORMATION SHEET

You are being invited to take part in research on developing a system for automated classification of common pigmented skin lesions for the general audience. Vinayak Sareen, BSc. Computer Science Student at Coventry University is leading this research. Before you decide to take part, it is important you understand why the research is being conducted and what it will involve. Please take time to read the following information carefully.

What is the purpose of the study?

The purpose of the study is to develop and investigate the reliability automated systems to classify the pigmented skin lesions which can be used by the general audience. The research will compare the classification results observed by medical professionals on pigmented skin lesions and the automated system to understand its reliability.

Why have I been chosen to take part?

You are invited to participate in this study because your input can contribute towards the research to investigate the overall reliability of the automated system for the classification of common pigmented skin lesions. The research will focus on investigating common categories of common pigmented skin lesions (melanoma, benign keratosis, melanocytic nevi, basal cell carcinoma).

What are the benefits of taking part?

By sharing your experiences with us, you will be helping Vinayak Sareen and Coventry University to better understand the medical performance of AI-based automated systems for common pigmented skin lesion detection. expectations of a medical professional from such an automated system and related concerns. Your data will be analysed and compared with the current performance of automated systems.

Are there any risks associated with taking part?

This study has been reviewed and approved through Coventry University's formal research ethics procedure. There are no significant risks associated with participation.

Do I have to take part?

No – it is entirely up to you. If you do decide to take part, please keep this Information Sheet and complete the Informed Consent Form to show that you understand your rights in relation to the research and that you are happy to participate. Please note down your participant number (which is on the Consent Form) and provide this to the lead researcher if you seek to withdraw from the study at a later date. You are free to withdraw your information from the project data set at any time until the data is destroyed on 17/04/2020 /until the data is fully anonymised in our records. You should note that your data may be used in the production of formal research outputs (e.g. journal articles, conference papers, theses and reports) prior to this date and so you are advised to contact the university at the earliest opportunity should you wish to withdraw from the study. To withdraw, please contact the lead researcher (contact details are provided below). Please also contact the Research Support Office [email esx072@coventry.ac.uk] so that your request can be dealt with promptly in the event of the lead researcher's absence. You do not need to give a reason. A decision to withdraw, or not to take part, will not affect you in any way.

What will happen if I decide to take part?

You will be asked a number of questions regarding common pigmented skin lesions and it's current diagnostic methods. The questionnaire/interview/focus group will take place in a safe environment at a time that is convenient to you. Ideally, we would like to audio record your responses (and will require your consent for this), so the location should be in a fairly quiet area.

BIBLIOGRAPHY

Meeting Logs

The overall structure and phases of conducting the research project was explained in the session. Based on my interest to develop cancer detection automated system my supervisor suggested to find appropriate dataset for training neural network for classification project.	Sep 17, 2019	EC1-11	13:00-14:00	Brief Introduction to the Project Structure.	
The referencing management softwares was suggested for bibliography and citation work such as mendley or reworks. In addition it was recommended to perform the personal research which will help to develop neural network throughout the research project.	Dec 12, 2019	EC1-23	11:00-12:00	Referencing Management Software and Project Research	
In this meeting the progress on the personal research and primary research plan was discussed with the supervisor. To ensure the ethics of the research supervisor guided to conduct the research in gdpr compliant manner and completing the form of ethics application before beginning the primary research.	Jan 22, 2020	EC1-01	13:00-14:00	Project Proposal Meeting and Ethics Application.	
This meeting was held at the programming support center. The supervisor guided me about the design and structure of the overall questionnaire. Supervisor guided me that participants should be able to select multiple options for each question.	Feb 05, 2020	PSC	14:00-15:00	Questionnaire Design	
+ Supervisor checked that the citations for the project was correct. However, there were some issues with the formatting of the urls in the report. Structure of reflection and conclusion section was also explained.	Mar 20, 2020	Online	12:00-13:00	Review, Guidance on Reflective Conclusion Section 	
Methodology section was missing the accuracy of the trained models which was required to be included. Furthermore, for the evaluation section guidance from sigma maths support center was	Mar 27, 2020	Online	13:00-14:00	Methodology section requires accuracy of each model. Testing and Evaluation	