

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](http://www.coventry.ac.uk/ipr) or contact [ipr@coventry.ac.uk](mailto: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 listed below

Sign: **Vinayak Sareen.**

Date: **07/03/20**

# Abstract

Abstract should be a succinct and self-standing summary of the basis, context and achievements of the project. Minimally an abstract does three things: (1) It states the problem that you set out to solve, (2) It describes your solution and method, (3) It states a conclusion about the success of the solution. Be straightforward and factual and avoid vague statements, confusing details and "hype". Do not be tempted to use acronyms or jargon to keep within the half-page limit. Consider that search engines, librarians and non-computer scientists wishing to classify your Report rely on the abstract. You may if you wish provide a short list of keywords (2-6 is reasonable) at the end of the abstract.

# Contents

# Acknowledgements

I want to express my gratitude to my supervisor Dr David Croft for guiding me through the investigations. The support of all medical professionals who contribute by providing data to the research is sincerely acknowledged. At last, I would thank my parents, who supported me financially and psychologically through 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 tumour caused by melanocytes can result in uncontrolled and abnormal growth which can spread in the human body (?). Detection and classification of unknown pigmented skin lesions can result in early diagnosis of the medical problem. The research data provided by 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 (?). Dermoscopy is a non-invasive method of examining the pigmented skin, which includes microscopic imaging of the surface structure of pigmented skin lesions (?). 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 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 (?). The current diagnosis method of detection involves using ABCD rule which considers the Asymmetry, Border irregularity, Colour irregularities, Dermoscopic structures respectively of common pigmented skin lesions (?). 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 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 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.

## 1.2 Objective of Research

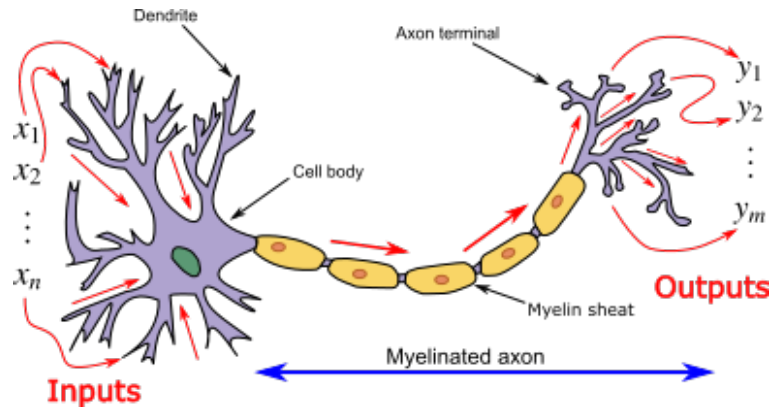
The research concentrates on developing a type of artificial neural network called a convolutional network to perform automated optimal analysis to identify the class of pigmented skin lesion. The research focuses on providing the quantitative analysis on comparing the results predicted by the automated 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 analysing limited skin tumours such as melanoma, benign keratosis, melanocytic nevi and basal cell carcinoma. The hypothesis for the research is that model will perform better than medical professional in examining the pigmented skin lesions. The investigation employs publicly available HAM10,000 dataset (?). The HAM10,000 dataset contains extensive collection of 10,000 images of labelled data units were collected from a diverse population of subjects over twenty years. The outcome of the research project will to be analyse the effectiveness of the automated pigmented lesion system. Furthermore, the intelligent model will be deployed on web based system to provide interface to use the system to analyse by general audience.

## 1.3 Perceptron Model

### 1.3.1 Inspiration from biological Structure of Neurons

The human brain are composed of millions 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 (?). The inspiration of artificial neural networks are taken from functioning of human brain and learn to recognise patterns and relationship in the data (?).

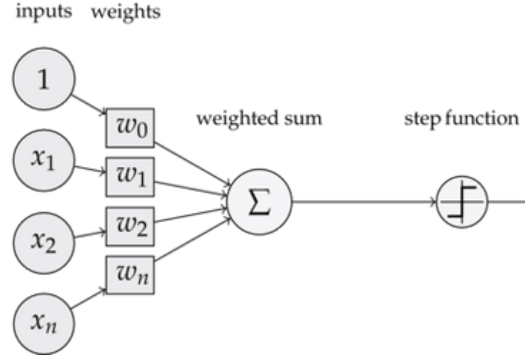
### 1.3.2 Neural Structure



The figure above represents the biological structure of neurons that helps in the communication of electric signals in the nervous system to learn and process information. Biological structure of neurons includes three major 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 neuron in the human nervous system through axon terminals (?).



#### 1.3.3 Perceptron Model



Preceptron model was proposed by Frank Rosenblatt in 1943 to design the model to mimic the human brain (?). Preceptron model or single layered feed forwards network networks takes the vectors of inputs and multiply with a randomly initialised weights and add random bias to network and process the information by providing data to the activation function to process the information (?). Figure above represents the simple preceptron model which consists inputs  $x$  and weights  $w$  for which weighted sum of multiplication result of inputs and weights will be passed to step activation function.

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

## 1.4 Multi Layered Feed Forward Neural Network

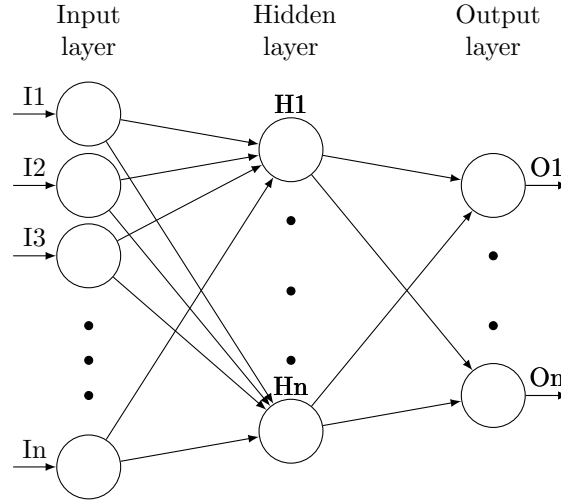


Figure 1.1: Multi Layered Neural Network diagram.

The perceptron model acts the base for all the functioning of modern multi-layered neural networks. The output of the single neuron described in the the above perceptron model is taken as the input in next layer of network. The above figure(1.1) shows an example of multi-layered neural where the first layer is known as input layer and intermediate layers are called hidden layer and last layer is known as output layer. The single neuron can performe limited computation but the computation power increases with inter-connected neurons in the network (?) where at each layer information is processed through the appropriate activation function. Artifical 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 (?). The training data is passed to the neural network to recognise the patterns in data and in each iteration or epoch the model predection gets improved with the optimisation algorithm which propagtes through the network to update the weights and bias to increase the accuracy (?). 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 model is trained. The fundamental rule for learning in artifical model

### 1.4.1 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 (?).

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

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 (?). The algorithm which aims to minimise the cost function is gradient descent. Gradient descent algorithm is a optimisation algorithm which is iterative in nature.

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

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

The algorithm functions by finding the partial deravative of the cost function in respect to weights and bias as shown in the equation above. The objective is to find minima of the cost fuction which improves the accuracy (?). Backpropagation utilises the gradient descent algorithm and propgates 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 (?). 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 (?). The ugly duckling signs state the spot which is unlike other lesions are great suspects of melanoma (?). Despite the limitations of above methods, both provide a great framework for clinicians and general audiences to spot melanoma based symptoms (?). 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 (?). Dermatoscopy is non-invasive microscopic imaging of pigmented skin lesions which provides clear imaging to perform proper analysis on pigmented skin lesions (?). 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 (?). 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 (?). 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 (?).

### **2.0.2 Support Vector Based Machine**

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 (?). 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.

### **2.0.3 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 (?). 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.

---

## 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 (?). 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

The research purposes a solution based on deep convolutional neural networks with various experiments mentioned in further sections. The operations require machine learning specific configuration for Cuda libraries and configuration to run programs on available GPU(Graphical Processing Unit) for faster processing and further, instructions to setup environment can be found at <https://www.tensorflow.org/install/gpu>. In addition, 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. The data science and machine learning libraries such as NumPy for numerical operations, matplotlib for visualisation, Keras for developing deep learning models and OpenCV for image processing was used in the process of developing an automated system for classifying pigmented skin lesions. HAM,1000 dataset was used to train image classifier, which can be found at <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T>. The dataset contains two folders containing dermatoscopic images of pigmented skin lesion and CSV file which contains meta information and image name for each pigmented skin lesion.

### 3.1 Data Processing and Normalisation

Datasets contain unclear and hairy images of pigmented skin lesions which were manually removed from the dataset to enhance the quality of available data. The information was read using pandas into the data frame, which is a data structure that allows storing tabular data from CSV files. The CSV file contained irrelevant information such as gender and age of patients and unbalanced data classes data columns were dropped from the dataset. Furthermore, 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 to preparing the dataset was reading the images data into NumPy array for both training and

### 3.1. DATA PROCESSING AND NORMALISATION

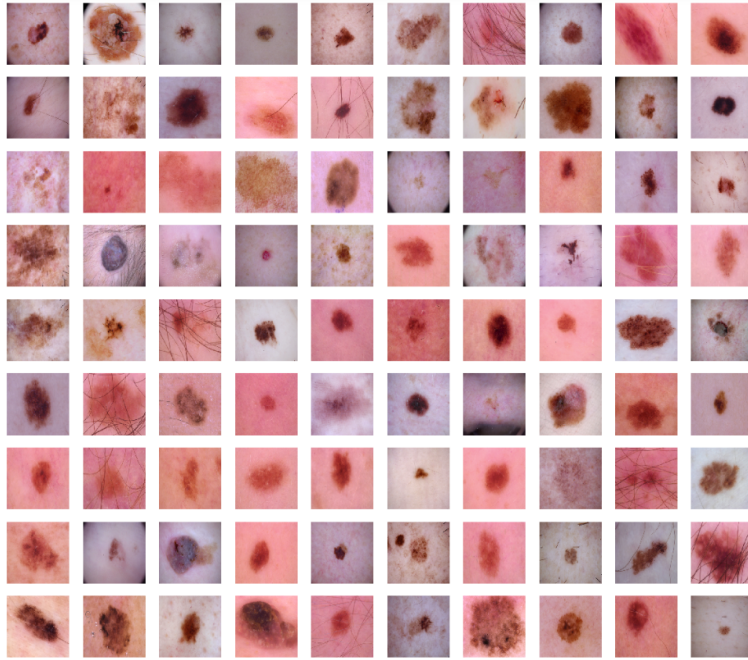
---

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.

#### 3.1.1 Data Normalisation

The images RGB images numpy array was mutli-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 categorical variable as binary vector and normalise the categorical labels into binary vector.

#### 3.1.2 Image Segmentation



The figure above reflects the sample from training dataset before performing image segmentation. The image segmentation was performed on the all the images using binary thresholding in OpenCV framework.



## Chapter 4

# Evaluations and Results

In this chapter, you should evaluate what you have done, and say what answer (to your research question) you have arrived at. It may be that in your method you describe some experiments, and this section records your results and analysis of those results. This is an important section – most students gain or lose marks in either their literature review or evaluation. The key to producing a convincing evaluation is to plan very early in the project what information or results you will need to write this section.

## Chapter 5

# Project Managment

This is a discussion based chapter BLAH BLAH BLAH .... Your first supervisor may have a very good idea of how well you tackled your project - however second supervisors may not have any idea. For this reason you need to include an account of the conduct of the project. What problems you encountered, how you overcame them, how diligently you worked, how you sought advice, how you responded to feedback. This chapter will be evidence driven – which is why you need to keep a log or diary of your project, maybe a project management timeline with milestones, keep evidence of each supervision meeting (signed off by your supervisor), Keep notes of supervisor feedback to your presentation and reflect on them in this chapter.

## Chapter 6

# Reflection

This is dummy text for reflection

## Chapter 7

# Conclusion

conclusion can be drawn from these examples to be continued .....



---

## Chapter 8

# Appendices

### Participation Consent Form

|                 |
|-----------------|
| 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 |

21

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

|                    |      |           |
|--------------------|------|-----------|
| Participant's Name | Date | Signature |
|                    |      |           |
| Researcher         | Date | Signature |
|                    |      |           |

---

# 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](mailto: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 its 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.

---

## Supervisor Meeting Logs