IOT DEVICE TO DETECT ANEMIA

Project ID: 19-129

Final Report

M.H.M. Akmal

M. Pravienth

N. Kayanthan

H.S.M.H. Fernando

B.Sc. (Hons) in Information Technology Specializing in Computer Systems and Network Engineering

Department of Information Systems Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

September 2019

IOT DEVICE TO DETECT ANEMIA

Project ID:19-129

Final Report

B.Sc. (Hons) in Information Technology Specializing in Computer Systems and Network Engineering

Department of Information Systems Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

September 2019

Declaration

We declare that this is our own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also, we hereby grant to Sri Lanka Institute of Information Technology the non-exclusive right to reproduce and distribute my dissertation in whole or part in print, electronic or other medium. We retain the right to use this content in whole or part in future works (such as articles or books).

Name	**Student ID**	**Signature**
M.H.M. Akmal	IT16134072	
M. Pravienth	IT16048324	
N. Kayanthan	IT16123014	
H.S.M.H. Fernando	IT16136120	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature of the supervisor: Date:

Abstract

Detection of diseases is a key aspect when we consider the world today. Diseases if not cured early it gets worse day by day. Therefore, early detection of the disease is a key aspect that needs to be considered. In this project the team shall be developing a device to detect Anemia which is a disease caused due to the lack of Fe3+ ions in the body. This is a disease that would even lead to organ failure eventually leading to a heart attack or death. Anemia is causing the red intensity of blood to reduce. Therefore, the team shall be using that factor to detect the disease. The team have built this device using a key symptom where the intensity of red in blood is measured and the output is sent to a central server, whilst the server also gets input from a questionnaire that is built in app. Using these data, the server will decide whether the user is suffering from anemia or not.

Acknowledgement

We would like to thank our supervisor Ms. Shahika Lokuliyana for the immense support and the motivation for our research project. Our supervisor also showed the path to carry on the research works and the discussions were very helpful to continue our work.

Finally, we would like to thank all the team members of the project for their contribution and effort towards achieving the project goals and objectives.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Description
GUI	Graphical User Interface
OS	Operating System
UI	User Interface
iOS	iPhone Operating System
EIS	Electrochemical Impedance Techniques
AWS	Amazon Web Service
RFID	Radio-Frequency Identification

Introduction

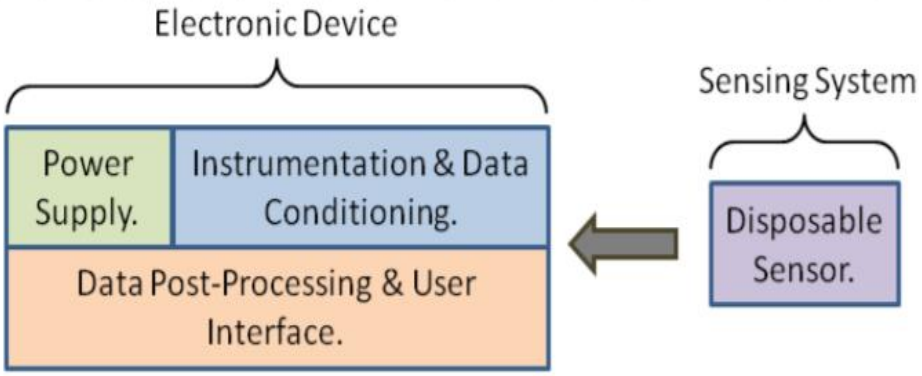
1.1 Background Literature

Anaemia, a disease that is mainly caused due to the reduction of iron in blood, which is a common blood disorder. If someone is carrying this disease, then the production of red blood cells in the body is low, which causes the protein known as haemoglobin to reduce. Due to the increment in different types of diseases we need better technologies for the identification of these diseases which the patient is diagnosed from. There was a research conducted by Jamie Punter and five other members. In this research they obtain 50µL whole blood sample to test for anaemia. The main aim of the project is to get a portable device so that it can be used for instantaneous detection of anaemia. The developed device contains electronic instrumentalization, post processing software and plug and play disposable sensor. The disposable sensor is based on a three-gold electrode commercial sensor which is of low cost and 50µL of blood is required for the device to use in this test. In order to achieve success with this device they had used 48 blood samples for testing of this device. These samples were collected from different clinics and hospitals for the task. Blood samples were distributed in two main groups as one for system calibration and the other 38 samples for system validation. The calibration of the device was done using a complete EIS experiment in this project to get accuracy in the detection of anaemia, defining the working range of haematocrit detection. So, in this project the specialty they had used an instant impedance detection in order to make sure they achieve accuracy, sensitivity and coefficient of variation in this project. After the proper testing has been carried out only 2% accuracy error had been seen [1]. Therefore, this is a more successful device that is portable.

But the main difference between the project that we are developing, and this project is that we will be also developing a portable device, but our project is non-invasive where we will not be getting any blood samples to carry on any tests but rather make sure that the required input is taken from the symptoms. For this we shall be building a device with the necessary features with image processing integrated into it. The next most important difference between the device that we are to develop, is that the device would be containing a connection to server which in our case in the AWS (Amazon Web Services) which is cloud which we shall be connected to send the relevant data and those would be connected to the mobile device of the patient. The patient should also answer a questionnaire that is designed in the app and the accuracy in ours would also increase

because we shall be accessing many symptoms and mainly using the device we shall be taking a main input which to determine whether the patient is diagnosed with this disease or not. We are going to detect whether the blood flow in the patient is low by using image processing and relevant technologies. Using these the team can detect the blood flow and determine the severity of anaemia in the body. This would be the idlest method to detect anaemia in a non-invasive methodology.

According to the above-mentioned project by Jamie Punter the structure of the internal



systems is as follows;

Figure 1.1.1: Schematic development view [1]

But the expanded view of the electrode sensor is shown below.

$$1 + j\omega C_M (R_1 + R_E)$$

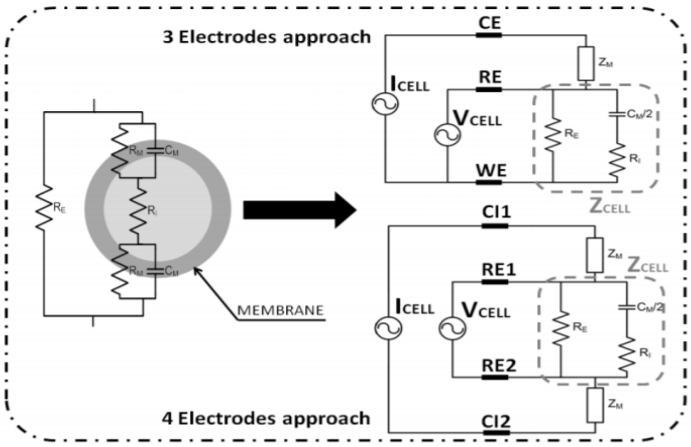


Figure 1.1.2: System based on 3 electrodes [1]

So, based on the figure 2.2 above figure the blood is dropped on the electrode and then the specific information that is gathered is transferred to that unit where it can detect the relevant changes and the testing happens.

The next project that is done by R. Bhattacharyya which is using RFID the ability to test for anaemia. In their project which is titled, “Towards low-cost, wireless blood anomaly sensing: An RFID-based anaemia detection sensor” they use a method there they would be able to send a certain frequency to detect if there are an ample amount of red blood cells in a given quantity of blood inside the body. Therefore, in order to detect that they have been shall be using a frequency and send it back and retrieve the relevant data and analyse whether there is a difference between the normal person’s blood and an anaemia infected patient. In an anaemia infected patient anyways there will be lower red blood cell count. Therefore in there project they have used a sensor which is capable of reliably differencing between blood having 20, 30, 40 and 50% red blood cells concentration by volume. In the RFID equipment that is used by this team has the ability of allowing for automated screening of blood specimens at large scale is also available. Therefore, according to this project the Red Blood count is taken by volume and is checked for anaemia. The structure of the frequency transferring device is shown below;

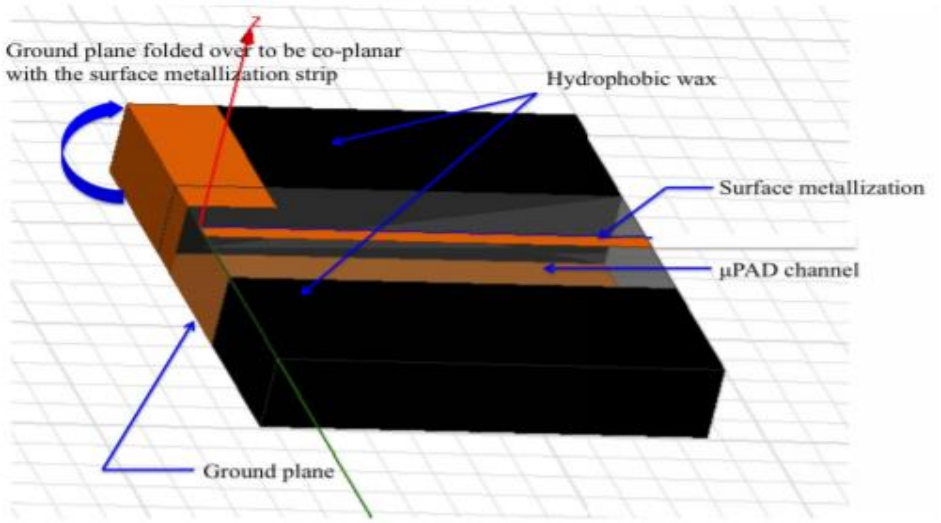


Figure 1.1.3: Micro strip designed with µ-PAD channel [2]

In the design there is a layer of wax so that the wave is not transferred to the required destination rather than making it to flow in a specific direction so that would facilitate the process of getting the inputs as well as outputs in an effective manner [2].

Image processing is also used in this project. Image processing is used in order to detect many kinds of changes, so in anaemia image processing can also be used to detect the exact symptoms and those changes. In a project done by Sachin D. Khirade to detect diseases in plants he has used image processing. In this scenario the researcher has been using steps like image acquisition, image pre-processing, feature extraction and classification to detect the changes in the plants by continually monitoring the changes in those plants by means of image processing.

So, in this project the image is captured initially, so at the capturing stage the image would be at RGB format so then the transformation for the leaf image captured would be changed according to the RGB colour transformations. Then image pre-processing step takes place so at that stage the noise that is there in the image is removed using different image pre-processing techniques like image clipping and etc. Then a histogram equalization is used in order to enhance plant image diseases. Then the image segmentation step takes place where the image would be portioned into various parts. So, in this step the segments would be portioned into different similar segments or having similar features. So, this could be done using different methods like otsu' method, K-means clustering a different method. Then this image is subjected to feature extraction. The feature extraction method plays an important role in identification of the

object. This stage is used in image processing to detect colour, texture, morphology and edges. So, in this project the texture is detected in order to make sure there are certain changes that could be detected. For this there are two methods that can be used. The first method is colour co-occurrence method and the other is leaf colour extraction using H and B components. Each of these methods can be used for getting a proper feature extraction. The next and the final step that is used is the classification step in this project for image processing. This is done using two methods, out of the two one is using the ANN method and the other is back propagation.

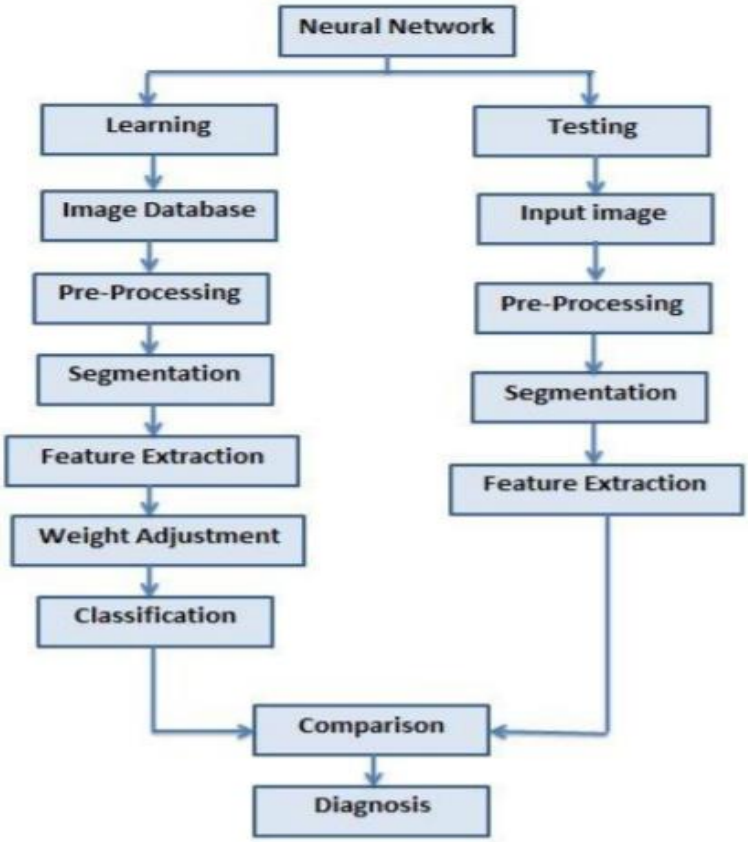


Figure 1.1.4: Working principle of ANN [3]

In the above-mentioned step-in figure 2.4 initially the image is captured and step by step learning process is run because there should be an image database that is created and then linked run through each process that can be seen on the steps indicated in the left hand side. While, when the device is implemented and is used the device would capture an image and that would be used and has to go through the testing stage, input image, pre-processing stage, segmentation and feature extraction, then the image that is produced would be compared with the images that was taken when the plant was at a

healthy state would be compared to whether there is a disease it has carried or not [3].

In the proposed project for the image processing the team would also be following some of the above steps to build the project. In the detection where would be taking images of anaemia infected patient and a normal person would be compared at the last stage after going through certain processes that are to be built. So, the images when the finger is pressed would be taken and processed using some of the method under image processing and then the image after relaxing the finger would also be taken and fed into the image database that is to be built. Then the image would be used for comparison afterwards.

CAD or also known as computer aided diagnosis is a major field that is growing in the current world. In order to achieve this machine learning techniques shall be used based on the systems that is to be introduced. In those devices the main feature would be collecting plenty of data in order to process them and determine the diagnosis for the disease-based on the symptoms that the patient carries. This can be done through analysing each symptom or through input from the relevant patient.

According the project by Meherwar Fatima on the topic, "Survey of Machine Learning Algorithms for Disease Diagnostic" has discussed on the types of machine learning algorithms that are available such as;

- Supervised learning
- Unsupervised learning
- Semi-supervised learning
- Reinforcement learning
- Evolutionary learning
- Deep learning

Therefore, these above-mentioned methods can be used in machine learning to detect the diseases. Pattern recognition and data classification methodology should be used to detect the exact cause and could be designed easily to detect the disease and make sure the patients would be able to diagnose the disease through these systems. Using different algorithms heart diseases, diabetics, liver diseases, dengue and hepatitis can be identified. For these proper algorithms should be used [4]. In our project the team shall be using machine learning algorithms in order to detect anaemia and increase the accuracy, which is crucial.

1.2 Research Gap

Prevention is better than cure, but early detection can get the patient suffering to the best cure. Therefore, identifying anaemia at the initial stage and taking necessary treatments would make the condition reduce and stop from reaching severe conditions. Anaemia can be identified from specific symptoms that are mentioned below;

1. General Fatigue
2. Dizziness
3. Pale Skin
4. Difficulty in concentration
5. Leg cramps
6. Insomnia
7. Shortness in breathing and headache, when exercising
8. Unusually rapid heartbeat
9. Cold feet and hands
10. Tongue swelling or soreness
11. Feeling faintish and blackout

So, based on the above-mentioned symptoms [5], identifying the disease that the patient is having is either anaemia or not is a key part of this research. Although there are many methods of identifying this disease most of them are done by laboratory tests and are more invasive methods, by taking blood and testing those. Which would take more time and needs complex systems to function to identify whether the patient is either diagnosed with this disease or not. Therefore, identifying the disease with a more portable machine is the best way to detect the disease. But the device would only be eligible for detecting one symptom but in the mentioned method the research group shall be using plenty of input to decide whether the person is diagnosed with this disease or not.

In order to do this, the team shall be using a non-invasive method. For this the team shall be developing a device and an app so that the team can get many inputs so that our accuracy increases when the team can state whether the patient is diagnosed with this disease or not. For that the team shall also come up with a device so that it can detect a certain symptom from the above-mentioned symptoms and finalize on a certain symptom so that the patient can have a double check with a blood check if needed and it would also show the same result.

This is a problem faced by women at large and the solution would also be helpful for many pregnant women for them to detect if they are diagnosed with this disease or not especially because the disease would be infecting the baby that they are carrying. Which would literally cause malnutrition for the child as the mother would be having lower oxygen levels. Therefore, if a solution is developed so that this can be detected at an early stage the patient can go the necessary treatments and the life of the patient and the child can be saved. Therefore, developing this device is important and the research group shall be implementing the device so that not only it will be portable but also transferring all necessary information through the cloud to the phone, which is the research gap and could be further improved if the historical data can be stored and be viewed later by the patient. This is the main gap that would be needed to be fulfilled. Although some attempts were made to make a portable device to detect anaemia there is no device to detect anaemia using the cloud and through multiple symptoms.

1.3 Research Problem

Anaemia, a disease that is mainly caused due to the reduction of iron in blood, which is a common blood disorder. If someone is carrying this disease, then the production of red blood cells in the body is low, which causes the protein known as haemoglobin to reduce. Reduction of haemoglobin would cause a major issue in the body, where it would reduce the amount of oxygen that would be carried in the blood would reduce. The reduction of this is a major issue because if the required amount of oxygen is not transferred to the organs then the organs would start to malfunction and sometimes organ failure could even be a possible outcome. This is like your body is starting to suffocate from within the body due to the lack of oxygen. This disease can be identified at the initial stage when a full blood count is taken if the amount of red blood cells is less than 12 grams per decilitre for women and 15 grams per decilitre the doctor would direct you for further testing because normally this reduction is caused due to anaemic conditions in the body. But to narrow down into the root cause the medical officer would direct those patients for further testing. Some of the identification symptoms that can be seen in an anaemic patient is that the patient would be having trouble in breathing. This is mainly caused due to the lack of red blood cells to carry oxygen, so the patient would develop heavily breathing conditions [6].

There are mainly 3 causes of Anaemia. The first is blood loss. When a person suffers from greater blood loss then a large amount of red blood cells is lost. During that time the

production is not enough to cater the amount of red blood cells the body needs. The probability of women getting anaemia is high due to the blood loss during the menstrual cycle. Heavy menstrual cycle periods would cause larger blood loss making women more likely to carry this disease. Other types of blood loss could be caused due to the amount of blood that is lost during accidents and external bleeding due to surgery or due internal bleeding that is very hard to identify. The second cause of anaemia is due to the reduction in the production of red blood cells. This condition can be caused by an iron lacking diet, medical condition such as cancer, AIDS or women who are pregnant also would be also anaemic due lower production of red blood cells, children also can be anomic due to genetic disorders of not producing enough red blood cells. The last cause of anaemia is the increased amount of red blood cells destruction by the spleen. This can be caused due to certain diseases that would cause our own body to destroy more red blood cells from our body. The second reason for this to happen would be an enlarged spleen that is causing the destruction of red blood cells that would cause anaemic conditions in the patients [7]. But there could also be other reasons like inability to absorb iron which a vital disorder that could last in anaemia soon, but this is a rare cause but the above mentioned three are the main causes of this disorder. So, identifying the disease early is the best solution for patients suffering from this problem.

According to statistics obtained from the WHO, globally, anaemia affects 1.62 billion people (95% CI: 1.50–1.74 billion), which corresponds to 24.8% of the population (95% CI: 22.9–26.7%). This is a large amount of anaemia patients that has been recorded. The highest amount is seen in children, mainly in pre-school children. The statistics of WHO is mentioned below [8].

Population group	Prevalence of anaemia		Population affected	
	Percent	95% CI	Number (millions)	95% CI
Preschool-age children	47.4	45.7-49.1	293	283-303
School-age children	25.4	19.9-30.9	305	238-371
Pregnant women	41.8	39.9-43.8	56	54-59
Non-pregnant women	30.2	28.7-31.6	468	446-491
Men	12.7	8.6-16.9	260	175-345
Elderly	23.9	18.3-29.4	164	126-202
Total population	**24.8**	**22.9-26.7**	**1620**	**1500-1740**

Table 1.1.1: Worldwide prevalence of anaemia [8]

In Sri Lanka anaemia can also be detected in high amounts. In Sri Lanka the prevalence of anaemia is high among women. This can be seen from various surveys that was carried out in Sri Lanka. According to indexmundi, “Prevalence of anaemia among women of reproductive age (% of women ages 15-49) in Sri Lanka was 32.60 as of 2016. Its highest value over the past 26 years was 48.70 in 1990, while its lowest value was 30.10 in 2010” [9]. According to a report by the government of Sri Lanka, which was published in the year 2006/07 it shows a report where the anaemic condition in school children also shows a calestropic rate. Mild anaemic conditions can be seen in many children which goes undetected. The statistics are shown below.

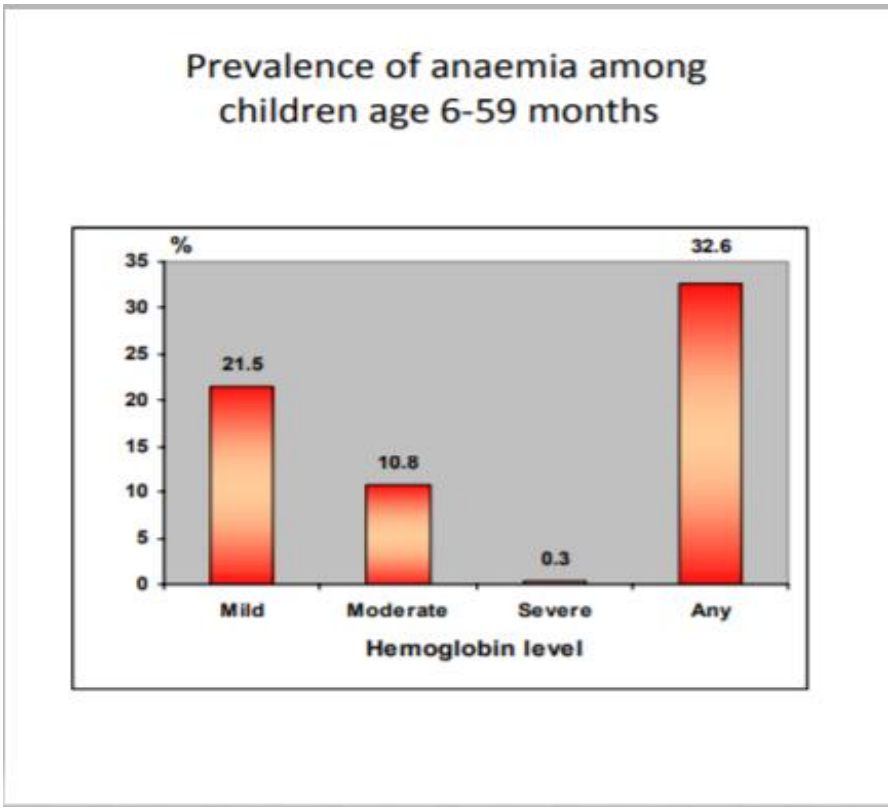


Figure 1.3.1: Prevalence of anemia among children

Anaemia is a disease mainly caused due to loss of blood or sickle cell, should be treated as soon as possible. For treatment to happen early detection should be possible before it reaches severe conditions which may turn catastrophic for the patient. Some of the symptoms that can be observed in the patient are fatigue and is vital to shortness of breath, or chest pain. So, to detect anaemia, the haemoglobin levels are accessed, and certain blood tests are carried out.

The above-mentioned process to detect anaemia is a time taking process so many people although they are infected by Anaemia goes undetected especially because there are no proper methods of measuring anaemia. In medicine it is well known fact that early detection leads to early cures. Therefore, the team is designing a device in order to detect anaemia without extracting blood.

1.4 Research Objectives

The part of this research is the implementation but before the implementation of this project the team needs to collect the necessary data to implement this device. From the above-mentioned symptoms under the topic 1.2, one of the key symptoms is skin going pale. So, the researchers will be developing a device to detect if the skin is pale and oxygenated blood flow is less so that, the skin is pale would indicate a major factor that anaemia and depending on the input from the questionnaire on the app we can detect whether the patient is suffering from anaemia or not. The questions in the questionnaire would mostly be related to the symptoms of anaemia which is a major factor in the detection anaemia is. So, in order to implement the hardware device, we need the image of the fingertip of the patient when it is squeezed as well as released. The time gap between the release after the squeezing should also be calculated.

The main reason to do this is when squeezing the finger tip of a person the fingertip turns yellow in colour, so we need a camera module configured in the respective module used which is raspberry pi where the camera module will be configured. Then the image during the squeezing of the fingertip will be taken as the first shot, then the image would be taken after the squeeze is released. Then using image processing we shall be checking the time taken and the comparison. Then we would need to gather data from an anaemia diagnosed patient in order to get the data when the finger is squeezed and released the time taken. So, comparatively we will be running an algorithm of an infected patient and a normal person's data to check if the patient using the device has anaemia or not then the information is sent to the server. The server would detect the input and we would request the patient to use the app to register before doing the above functions in the app so that the relevant information is taken, and an account is created in the app. Then the app would contain a questionnaire where the questions related to the disease is needed to be answered by the patient in order to get a proper accurate result, so based on that we can get the input into the server from the questionnaire as well as the device.

After getting the input from both the device and the questionnaire we shall be running a machine learning algorithm in order to detect whether the patient has been diagnosed with anaemia or not. After the algorithm is run the results will be pushed to the app where the final user would be able to see if he/she is diagnosed with anaemia or not. The main objective of this research is to make sure the patients would be able to detect anaemia at an early stage and take necessary treatments so that they can be cured at an early stage.

2. Methodology

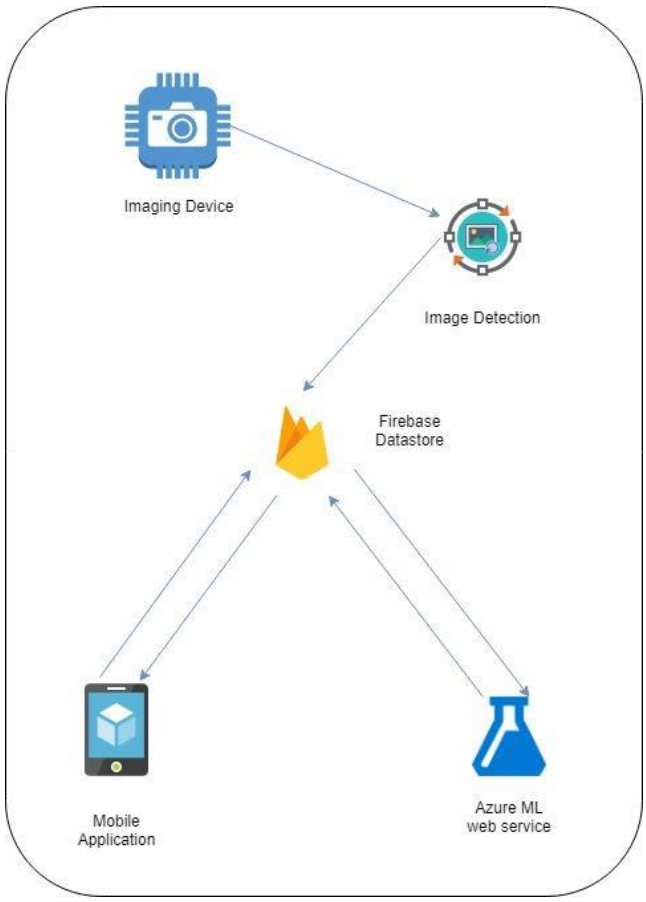


Figure 2.1: System architecture diagram

2.1 Procedures

The main procedures that is involved in the designing of this project is to build a simple device that has accuracy and give the user the ability to trust the results of the device. In order to achieve the above-mentioned objectives, the team have designed the project on 4 major steps.

- Building the hardware device

- Configuring Image processing
- Configuring connection between server and device
- Configuring the server and creating necessary channels to intake data from app
- Building the app
- Designing an algorithm in the server to process the data using machine learning algorithms

So, the above-mentioned steps are the key steps in building up this device and establishing connectivity in order to make the users able to detect whether they are diagnosed with anaemia or not.

2.2 Building the Hardware Device

Initially the device should be built so that the input can be sent from the device to the servers. Therefore, the team have decided in using a raspberry pi in order to configure the device. The raspberry pi would be connected to the camera module that will be used. The camera module would be designed in manner attached to a clipping device, where the device would be clipped on to the hand of the user and a picture of the squeezed finger would be taken. In order to take that the camera should be configured so that it is able to take the pictures in a more focused manner so that the level of identification is high. After the pictures are taken then they should be sent to the raspberry pi where the team shall implement image processing. So, when designing the device, the clipping should be done so that the finger is tightly held and then released upon the release during a specific period another snapshot of the finger should be taken. In this device a strong light would also be fixed so that it passes through the finger where it would make the blood visible for the camera and when the finger is tightened then the blood moves back and that also can be noticed. This would also include a snapshot when released and the time taken. Depending on that the images would be loaded in image database for image processing. The pictures would be eventually configured.

When the device is configured an OS will be used in the raspberry pi for the programs to run so that it would also facilitate image processing to happen in the same device where it would make it easy to configure image processing as well. The device shall

be connected to the server using node ESP or a similar device. Then the results that is processed in the device will be transferred to the server via the WiFi module of the Raspberry Pi that would be connected to the server. The TCP protocol will be configured, and it would be safe to transfer those details via this protocol because it is much safer than the other protocols.

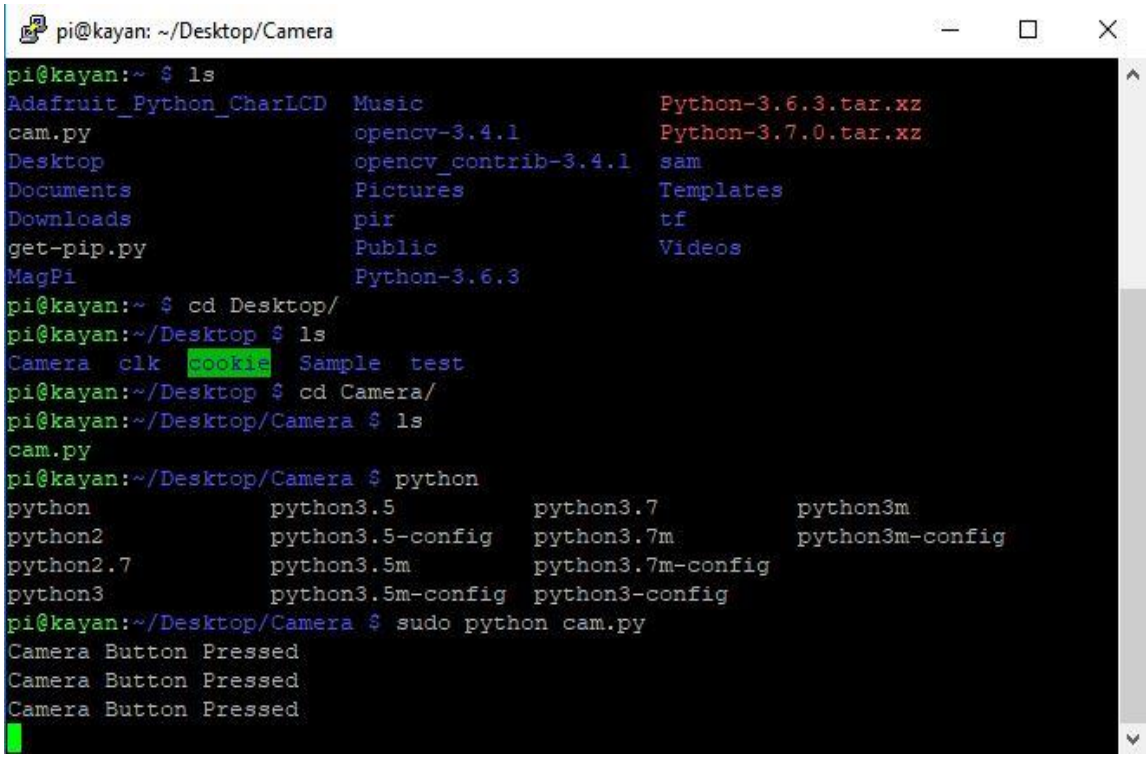


Figure 2.3.1: The camera being used

2.3 Configuring Image Processing

After the images are taken it would be sent to image processing at the same device. As the team have configured the raspberry pi with an OS the team would be able to make the necessary image processing the local server or in the cloud. So, when the images are sent there would be stored in a common location. One for the healthy patient and the other for the anemic patient should be classified when training the images.So, when configuring image processing the following steps should be considered;

- The input added to the processes
- Run pre-processing

- Segmentation
- Classification

Initially the images taken as samples should be inserted into the system. Then the images should be cleaned from noise to avoid it interfering in the decisions the device would arrive at the latter stage. Then at the pre-processing stage the images would be used for detection of the intensity of the images. The main aim of this step is to make sure an improvement in the image data and the unwanted information would be suppressed from unwanted distortions or enhances so that in the further processing part it would be giving the correct amount of data for decision making and the device needs to use two of the sample data for processing where this would apply for both the patient's image as well as the healthy person's fingertip image. In order to achieve this, the team can use many techniques.

The third step in image processing is to do segmentation. In segmentation the image is set to the process of portioning where the image is portioned into multiple segments such as set of pixels. The main aim of this step to make sure that the image is represented in a more meaningful manner so that it could be easier to analyse. This mostly used in order to determine the boundaries and objects that are in the image. This would help to clarify the exact shape of the image that is needed. The image processing steps are shown in the below diagram;

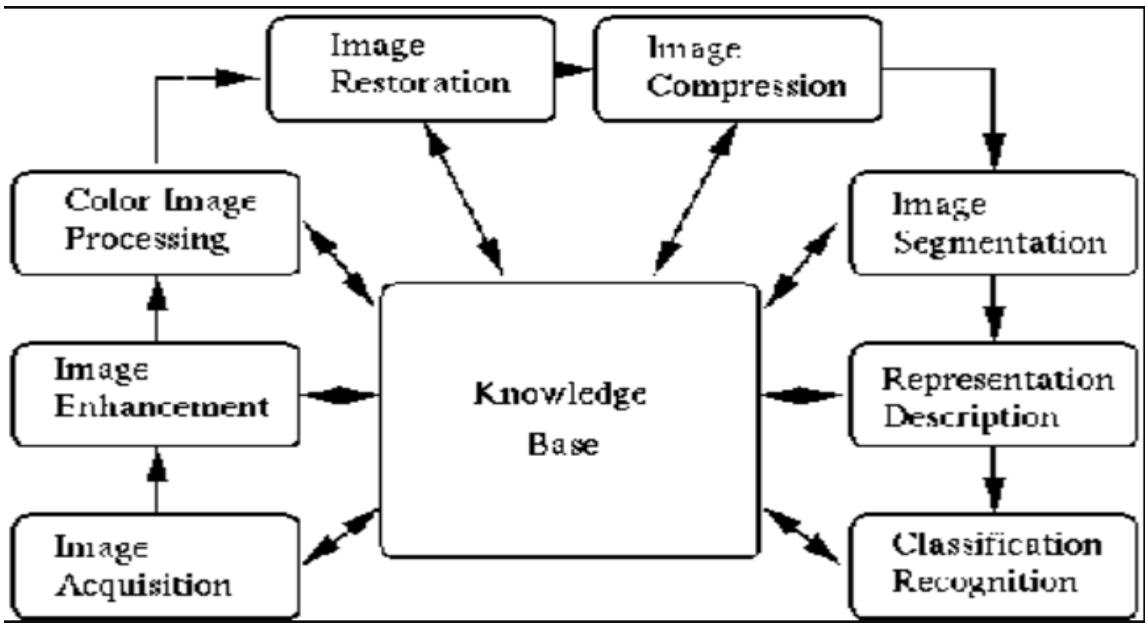


Figure 2.3.2: Basic steps of image processing [10]

The above flow of diagrams shows how the image processing steps work and each level has certain set of algorithms that is run to achieve the desired output. The next step after segmentation of the image is the step of classification. Although in figure it depicts representation description, the team shall be looking at the main steps that would be involved in the process of identification of the image. So, as the next main step, classification of the image, where the step categorizes detected objects into defined classes so that the image patterns can be compared with the target patterns. Although classification of objects by humans is simple task it's a complex task when it comes to computing. Because a quality image should be needed from the device a low-price high-quality camera and with automatic focusing capabilities should be used [11].

So, after this the recognition of the image would happen and then the data will be sent to the server in order for further processing. This would help to detect whether the user has anaemic conditions or not and the data which would be sent to the server would again subject to machine learning depending on the input given by the app as well as the device after image processing.

In the image processing component, Initially the image that is taken from the device is sent via Wi-Fi to the location specified in the image processing algorithm. Then the image is run against the model that has been trained and the output is sent to the firebase server that has been connected to the service. This python server shall transmit the relevant data to the collection in the firebase that has been defined and it sends the accuracy rate of the healthy patient. Then the service shall look for the next request to function similarly. This is the method to implement the cloud server service locally as this is ideal for the cost reduction.

Here we shall run a server instance locally created. A python server instance can be created and then the instance can be run. In order to run this python instance, we need a model of the images trained using the data set be available to use at any given instance. Initially the service should be able to fetch the image from the device after it has been taken and the image needs to be added to the service to be validated against the trained model that has been created using keras and tensor flow. Here the model is already trained using the sample datasets that have been collected using the anemia patient against the healthy patient's dataset. During the training of the data set we

have to classify the dataset accordingly according to the defining naming conventions for anemia and healthy patients in order to get proper trained models. In order to train the model, we shall be using a library called Imageai where the keras, OpenCV, NumPy, h5py, matplotlib and TensorFlow are integrated so that they should be pre-installed in the environment where the library is being used. Then after training this model the model is used to compare the image where the TensorFlow technology is used for the image comparison with the model where the service has been made available. Then the python server instance where the image recognition is configured the taken image from the device needs to be mapped.

The following images shown below in figure 2.3.2 and figure 2.3.3 are the images that are identified based on anaemia patient or not an anemic patient. The colour designation is identified here.

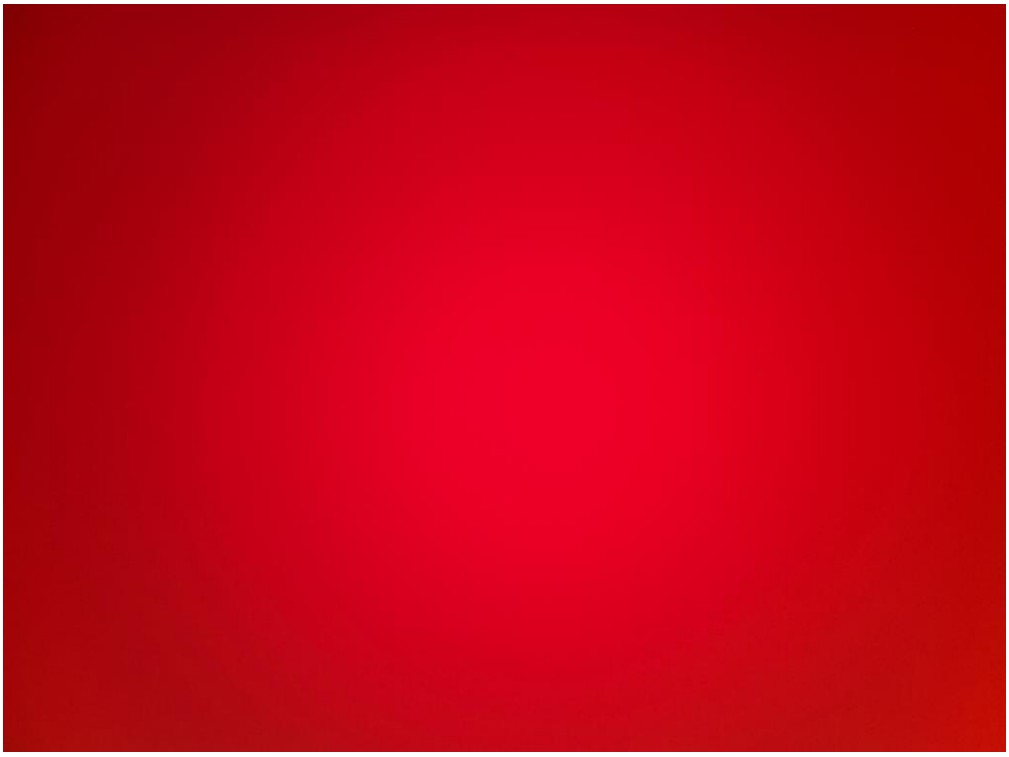


Figure 2.3.3: Data sample of healthy patient

Figure 2.3.4: Data sample of anemic patient

2.4 Configuring connection between Server and Device

In this step the connection between the server and the device should be established. In order to achieve this, the team need a WIFI – module or a module that would connect with the internet for transferring data to the server. In order to do that the team shall be using the WiFi module of the Raspberry Pi for the purpose. The device shall be connected to the WIFI by configuring the module using AT commands this would establish the connection between the server and device via API key that is given for each.

The Raspberry Pi device is configured with a python application which fire every time a photo is taken by the device. This application is configured with the necessary service account information needed to connect to the firebase datastore and send retrieve data from the datastore. This python script will seta flag in the datastore which will be used by the image processing application to process the image produced and store the result in the firebase datastore.

2.5 Configuring the Server and creating necessary Channels to intake data from the Application

The firebase datastore is connected to the mobile application by configuring the necessary information in the mobile application code which allows the application to read and store data in the datastore. A python server was setup to constantly monitor the datastore to detect when a user has uploaded data to be used in the machine learning web service, once detected to send the relevant data as a POST request to the web service and to store the result contained in the response to the POST request in the datastore. The portion of the server code used to compose the POST request sent to the machine learning web service and then to store the extracted result from the response to that request is shown in the figure below.

```
data =  {

        "Inputs": {

                "input1":
                {
                    "ColumnNames": ["Healthy", "Q1", "Q2", "Q3", "Q4", "Q5"],
                    "Values": [ [ IPR, Q1val, Q2val, Q3val, Q4val, Q5val ], [ "0", "0", "0", "0", "0", "0" ], ]
                },        },
            "GlobalParameters": {
}
    }

body = str.encode(json.dumps(data))

url = 'https://ussouthcentral.services.azureml.net/workspaces/c194828abb25492faafef1241411b535/services/5b3b31de418d45a0872cd8f433085cdf/execute?api-version
api_key = 'DTbhh434sD95RSAX/vlK/z70xgajuWTaCvVAFo6zqSq6Ej+mqQWgzY3Yic/jygNOQsn469JSvnEGuUh8XYuWQQ==' # Replace this with the API key for the web service
headers = {'Content-Type':'application/json', 'Authorization':('Bearer '+ api_key)}

req = urllib2.Request(url, body, headers)

try:
    response = urllib2.urlopen(req)

    # If you are using Python 3+, replace urllib2 with urllib.request in the above code:
    # req = urllib.request.Request(url, body, headers)
    # response = urllib.request.urlopen(req)

    result = response.read()
    print(result)
    print(result[131])
    score_label = result[131]
    score_label_value=int(score_label)
    print(score_label)
    score_probability = result[135]
    score_probability_value=int(score_probability)
    print(result[135])
    print(score_probability)

    doc_score_ref = store.collection(u'scored_predictions').document(u'IpHRjF8jqQnrC6Qj0md4')
    try:
        doc_score_ref.update({u'scored_label': score_label_value})

    except google.cloud.exceptions.NotFound:
        print(u'Missing data')
    try:
        doc_score_ref.update({u'scored_probability': score_probability_value})
```

Figure 2.5.1 : Python server code

2.6 Building the Mobile Application

After the device the next main mode of input is the app. The app would be designed so that there will be a user login so that the user will have to register before logging into the system. The user will have to use the email address to register to the system. Then according to the database, the system will be designed so that there are the users can use their username and passwords to access the accounts. When accessed the app using the proper credentials the user will have to two platforms, where in one the user will have to answer the questions that are designed in the questionnaire. In the app the questionnaire should be designed according to the symptoms of the disease. The symptoms as mentioned early would be;

- General Fatigue
- Dizziness
- Pale Skin
- Difficulty in concentration
- Leg cramps
- Insomnia
- Shortness in breathing and headache, when exercising
- Unusually rapid heartbeat
- Cold feet and hands
- Tongue swelling or soreness
- Feeling faintish and blackout [5]

In order to the disease, need to find the level of the disease with the help of application to detect the person who affected by Anemia or not. The mobile application helps to find the level of Anemia affected to the person. Earlier, in order to find the disease, they are using blood samples and crops red blood cells to detect anemia and it shows with the certain device or lab report. Our team designs the uniqueness to detect the disease by non-invasive method. It has no blood cells to check with the machine. It contains with the device that gets image processing and detects the anemia by mobile application.

The main aim of the research is detecting the anemia by image processing level and calculates the disease with the machine learning process. Then it gives the results

calculation with the questionnaire part. This is unique in the device and specially in the mobile application it shows the combination result of questionnaire and device results. This product gives best accuracy than the earlier products. Therefore, the product is increased their accuracy level in the product.

Features \ Products	Anemia detection Device by Jaime Punter(2015)	Anemia detection device by De Benoist B(2013)	Research product
Portability	✓	✓	✓
Non Invasive method	×	✓	✓
Number of symptoms accessed	Blood Test	Hb measurement	Based on 5 - 10 symptoms and image of fingertip
Mobile Application	×	×	✓
IoT implementation	×	×	✓

Table 2.6.1: Comparison between earlier devices and created device

The above table 2.6.1 differentiates between the earlier products of similar functionality and the product made by the team. The main features are added to the product of this project are mobile application and IOT implementation. It displays the separate results and saves the results of each patient. It is categorized with the patient and able to modify the results once a week or once a month with the help of related doctor.

The mobile application's objectives are the deep knowledge about the anemia symptoms to detect the anemia by mobile application. The mobile application connects with the data server with the connection IOT. It combines the patient's details, doctor details to get the results with the help of mobile application. The results given the best accuracy with the connection of questionnaire about symptoms and the results of the image processing. There is the option that directly connects the device to operate by the mobile application. The main part of the mobile application is the connection with IOT platform.

The mobile application is named as "Doc-Detector". It is an android based mobile application for Anemia analysis doctors. The application offers various operations

such as questions based on symptoms, details of Patients, viewing results of the disease level and the date of the next meeting session for the patient and doctors. These services are conveniently grouped and developed specifically for use on the Android device. All the operations can be performed when an internet connection is available. Administrator of the application can register the Doctor to the system to all services provided.

The mobile application has the separate database to save the data of the results, patients and doctors details. Most of the mobile application database connects with the local database and it has the SQL connection and structure to the database. In this project, the mobile application has the separate server to save the data within the database. The project is chosen the Firebase database because it is No SQL database structure and it is real-time database helps to store and synchronize the data. The database is fast and provides very secure hosting. The Firebase is Backend-as-a-Service (BaaS) that offers developers a wide spectrum of tools and services to develop high quality apps at a much faster pace [6]. This database can connect their applications with backend cloud storage and APIs rendered by the backend applications.

The methodology is given in the below flow chart diagram. It shows the flow chart of the doctor when login to the mobile application and it goes with the correct path of the system.

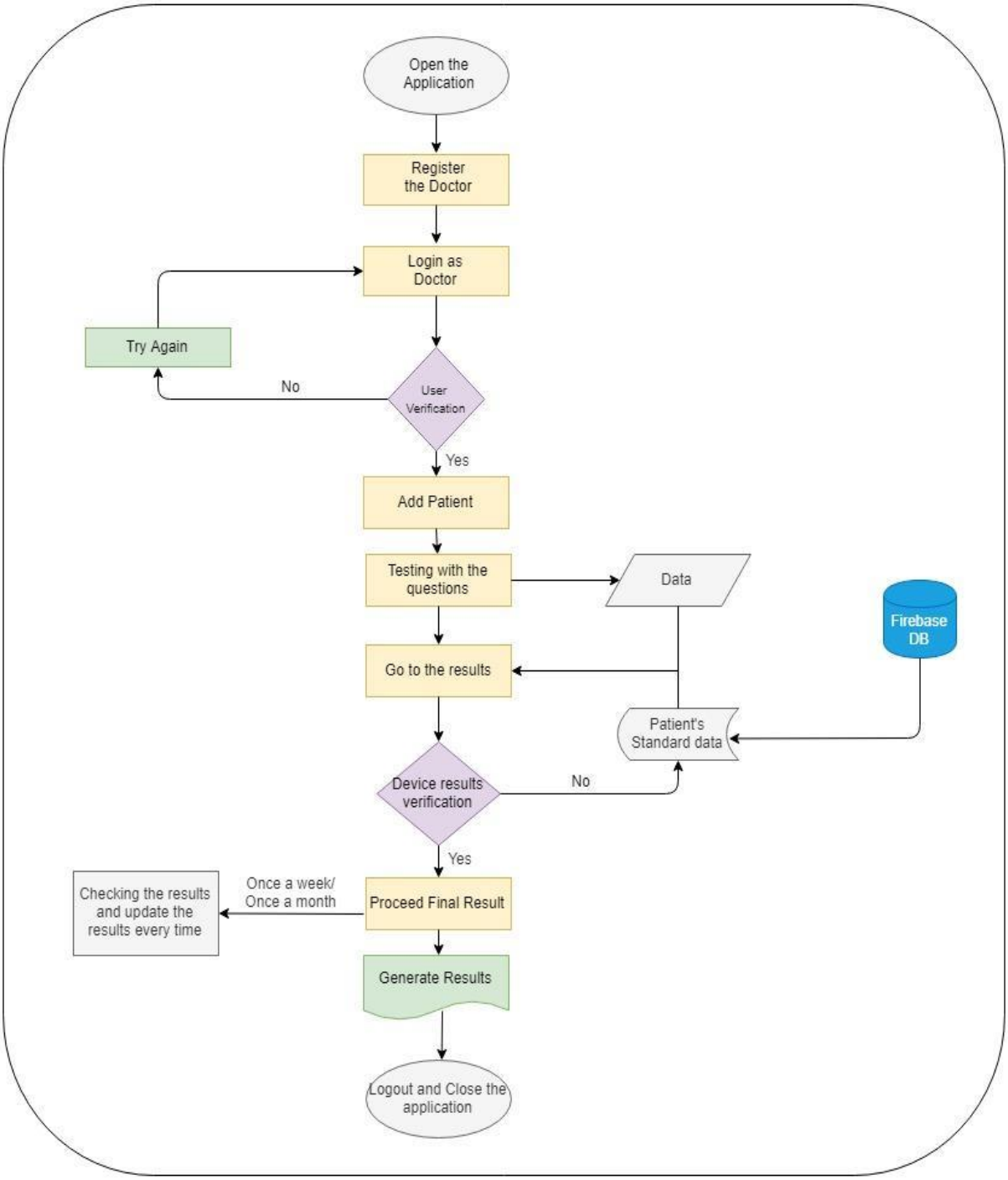


Figure 2.6.1: Flow Chart of the Application

The following external interfaces and functions are held in the mobile application in the procedure:

External Interfaces

- Welcome Page
- Login Page for Doctor and Admin
- Patients Info Page
- Patient Register Page
- Patient Personal Details
- Questionnaire Page
- Results Page
- Doctor Register Page
- Doctor Info Page
- Device Page

Functions

- Login – Admin
- Login – Doctor
- Add Doctor
- Edit Doctor
- Delete Doctor
- Add Patient
- Delete Patient
- Update Patient
- Logout

These interfaces and functions are designed in the mobile application to run the results and make the data clear to the doctor view. All processes are under Admin panel and it always directs to the Admin. According to the order of Interface and functions, it clearly helps to fulfil the requirements and fill in the blanks of the mobile application detected request to collect the data of doctor and patient level. The details can be captured in the back-end server to save the data in the server.

The mobile application is mainly to focus the patient's question basis and given the results section in the UI. The mobile application is created in the Android Studio Platform and it modifies for user friendly manner to use the application.

Figure 2.6.2: Shows "Doc-Detector" in Android mobile application

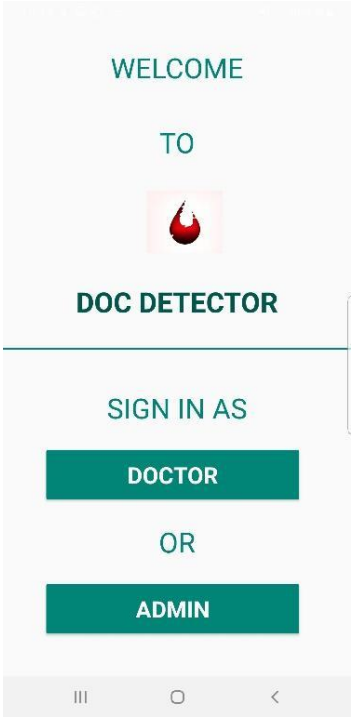


Figure 2.6.3: Welcome Page

The mobile application's database used as "Firebase" and it directly saves the question results and algorithm results. Then it calculates the total results and saves in the personal patient's database. Finally, it shows the results in the mobile application.

The implementation of Firebase data base is the component to connect all sectors in this project. The Firebase cloud store is directly connecting with the device which checks whether it captures to build the image processing scenario to take the snapshot. The database is the main segment of this project. It syncs the whole section of the team members. The mobile application questionnaire saves the results and the image processing saves directly in the database and machine learning data processing captures both data and predicts with the algorithm. Finally, the final results only save after the prediction of each patient category. The database needs authentication to show the data and it is able to use for selected users.

2.7 Designing an Algorithm in the Server to process the data using Machine Learning Algorithms

Several two class classification algorithms were compared using experiments in Azure Machine Learning Studio. The compared algorithms were boosted decision tree, boosted decision forest and boosted decision jungle. Experiments were run using a dataset that would be similar to our eventual dataset for this research with having 10 numeric valued attributes and with around 120 records. This is similar to the eventual dataset for this research so by running experiments on this dataset, the team was able to determine the best algorithm for this type of work by comparing the accuracy of each prediction model produced. The best algorithm turned out to be two class boosted decision jungle algorithm for this type of work.

The data from the image processing algorithm and the mobile application questionnaire must be collected and then transformed in various ways in order to build a data set that can yield an accurate predictive model when put through the learning algorithm. Qualitative data must be formatted to qualitative, numeric data and for each variable the weight of the variable must be determined so that a desirable prediction model is produced by the learning algorithm. Over 105 data samples were collected from which only 91 were used in the final dataset used to train the predictive model after applying data cleaning and transforming techniques. The final dataset was then composed in .csv format as shown in the figure below.

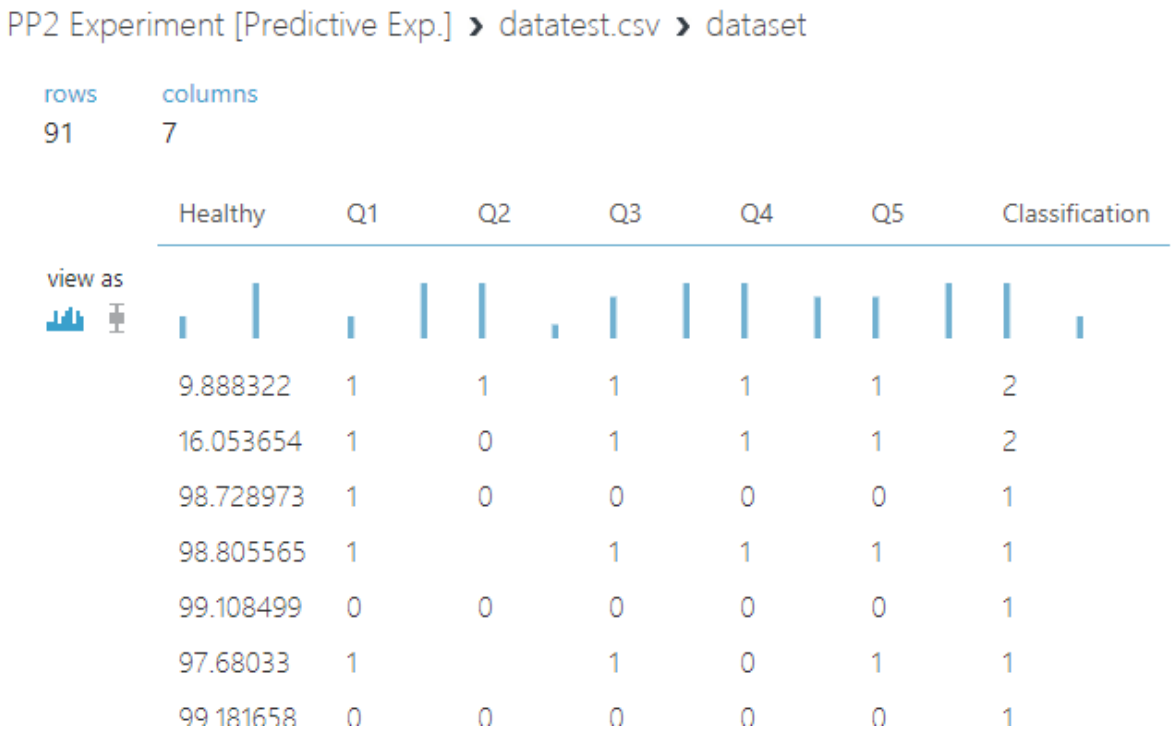


Figure 2.7.1: The dataset

As can be seen in the figure above 91 records were present in the dataset with 7 columns. The first column represents the healthy classification percentage generated from the image processing algorithm for each user. The next columns from Q1-Q5 represents the answers to the questionnaire provided by each respective user with 0 representing false and 1 representing 2. The final column labelled Classification is the actual anemic diagnosis for each user with 1 representing a healthy status and 2 representing anemic status.

After building the dataset, the prediction model must be trained until a prediction model with desirable accuracy levels is produced. The following figure shows the training completed for the prediction model.

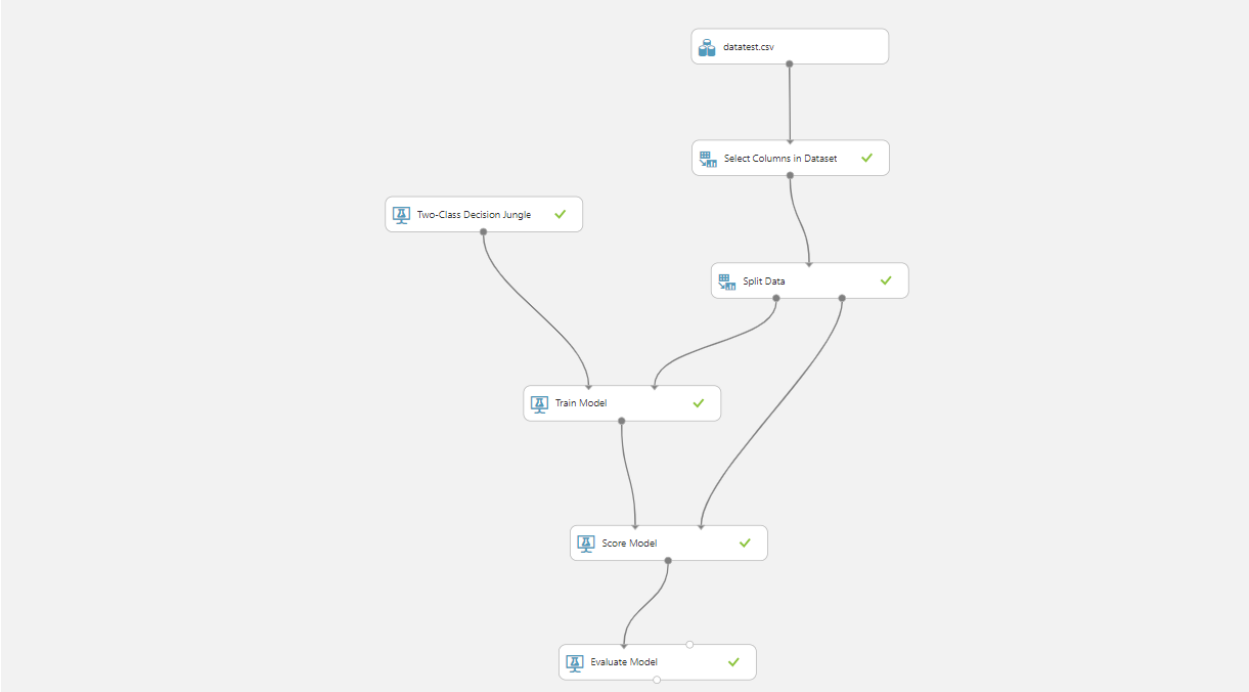


Figure 2.7.2 : Training

After obtaining an adequate prediction model it is then necessary to implement the prediction model as a web service so that the backend can send API requests and obtain results for patients. The following figure shows the implementation of the prediction models used as a web service through Azure Machine Learning Studio

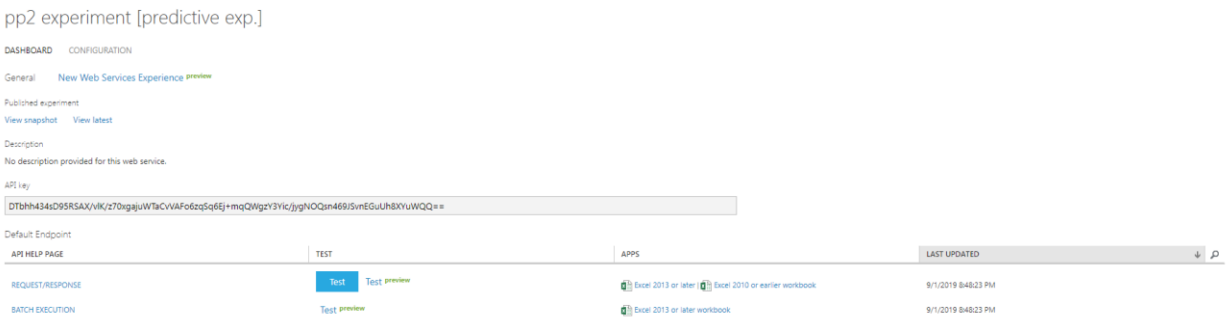


Figure 2.7.3 : Web service

As can be seen an API key is provided for each web service so that requests can be made to the service and obtain the results form the prediction model.

Testing will be carried out by using a portion of the dataset as shown in the figure below.

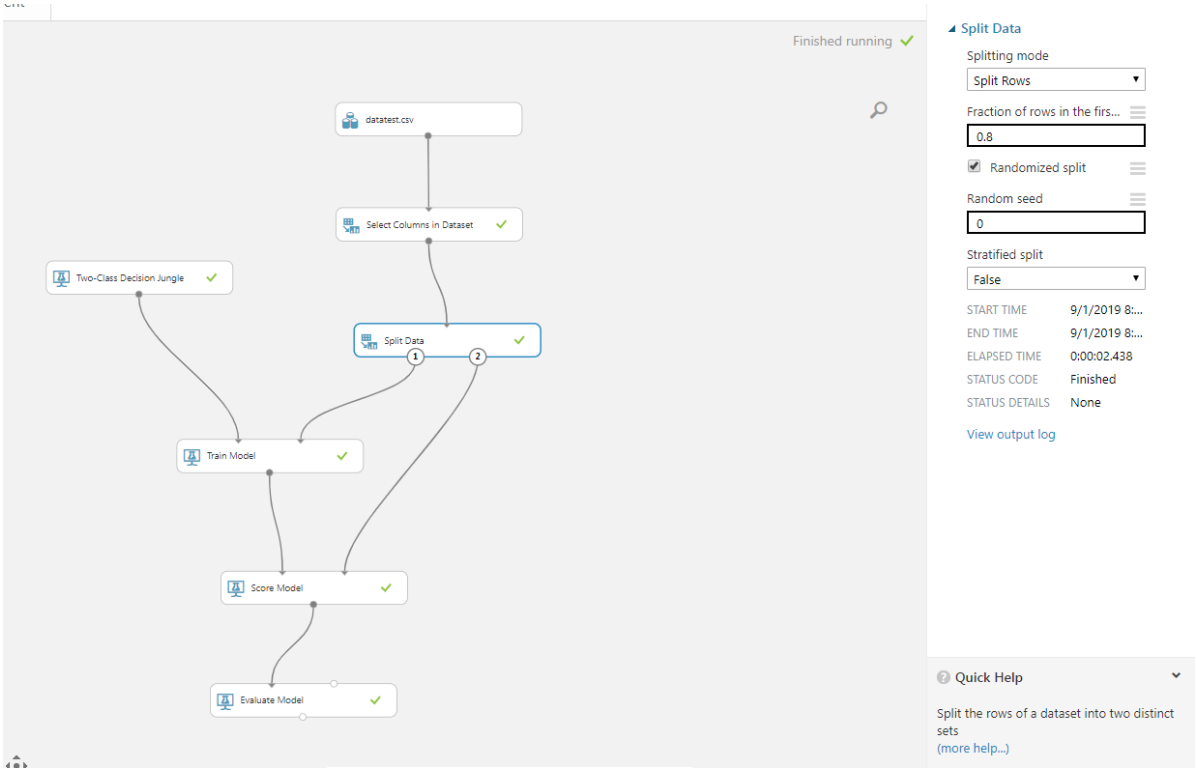


Figure 2.7.4 : Testing model

As shown in the image above the dataset will be split into two portions and one of the potions will be used for testing of the prediction model produced by consuming the other portion. 80% of the dataset is used to train the predictive model while the other 20% is reserved to test the accuracy of the model.

2.8 Commercialization aspects of the product

This project was built on a commercialization concept. The sugar machine although requires an invasive method it has been able to capture a larger market due it's accuracy rate. If the accuracy rate is high, then the accuracy of the product can be increased. In order to increase the accuracy rate of image processing the team has used the following concepts where the keras deep learning method is used where the accuracy rate is higher than usual method. The symptoms that is also taken during the construction of the device also needs to be taken into account when it comes to

the commercialization of the product. The device only inputs one symptom while the other component of the total system which this app inputs some information regarding the anemia patient is, where the input is taken from the questionnaire where the user has to answer the relevant questions. This also increases the accuracy of the device further in the process of detection of anemia.

As anemia is a widespread disease there is demand for a detection system especially for non-invasive systems. The device is designed to be used by many users. The service-based products used to implement the system means that the system can be scaled efficiently and effectively with increasing demand. Applications from other platforms such as web applications and iOS applications can also be connected to the same firebase project so this means that if the need arises due to commercial demands, those applications can be integrated very easily. Both Azure and Google offer reasonable service packages as the need to scale the system increases. Also with the experience and knowledge gained from this research, a mobile application that detects anemia using the phone camera can be developed which could be a huge commercial success.

2.9 Testing and Implementation

Testing should be done by corroborating the results with results taken from a sample of anemia patients. Accuracy of the diagnosis as well as other functionality must be ensured. First stage of testing will comprise of using the split data from the dataset to immediately determine the accuracy of the system. Then further testing must be done by using the system to test another sample of users and comparing the results against blood tests done for those same users. During testing there can be setbacks due to the bugs and errors. But the fixes should also be done as soon as possible and make sure the device is functional and also be a guarantee that the information that is obtained taken through the device is also kept securely because the medical information are sensitive information therefore the proper security needs to be provided. After full system integration, various testing phases were carried out to ensure the robustness of the system.

3. Results & Discussion

3.1 Results

After the implementation of the device the research team was able to test it on a certain number of individuals who were healthy as well as who were diagnosed with anemia. The results were initially showing an accuracy rate of as shown below. But with continuous training that the device was subjected to it was able to show an accuracy rate which is more than 80%. The device has then been configured with an external design so that it can be marketed at a later stage. When configuring the above-mentioned models there were certain problems which would be faced when configuring the image processing as the data samples which has been collected for the implementation of the demo is not exceeding more than 150. The amount of data sets that should be collected in order to train the system is manually done as the data set is not available therefore if a large sample of dataset can be collected then it could be used to train the system more accurately that would make the system eligible to attain a greater result with a better accuracy. But currently due to the low amount of data that is available a low amount of accuracy is arrived at. The ability to make the device portable and the usage of non-invasive method is an attractive feature that leads more users to suggest the device. This reduces the blood loss in anemia patient facilitating the users the device fearlessly. The data processing algorithm to train the predictive model which produces the results was set up in Microsoft Azure Machine Learning Studio as shown in the figure below.

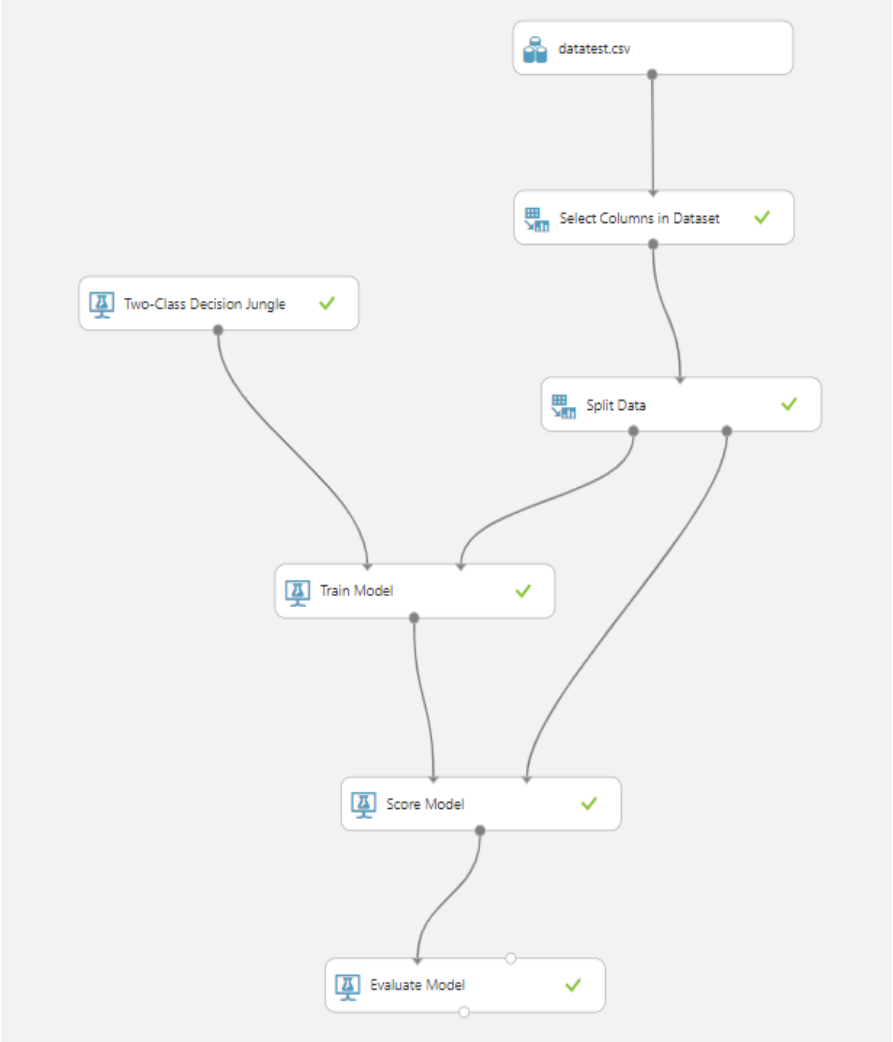


Figure 3.1.1: Predictive model structure

The dataset is split into two portions in 80:20 ratio and the first portion are used to train the model while the latter portion is reserved to test the produced predictive model in order to measure its accuracy.

The result of the evaluation is shown in the figure below with the predictive model achieving an accuracy of 82.6%.

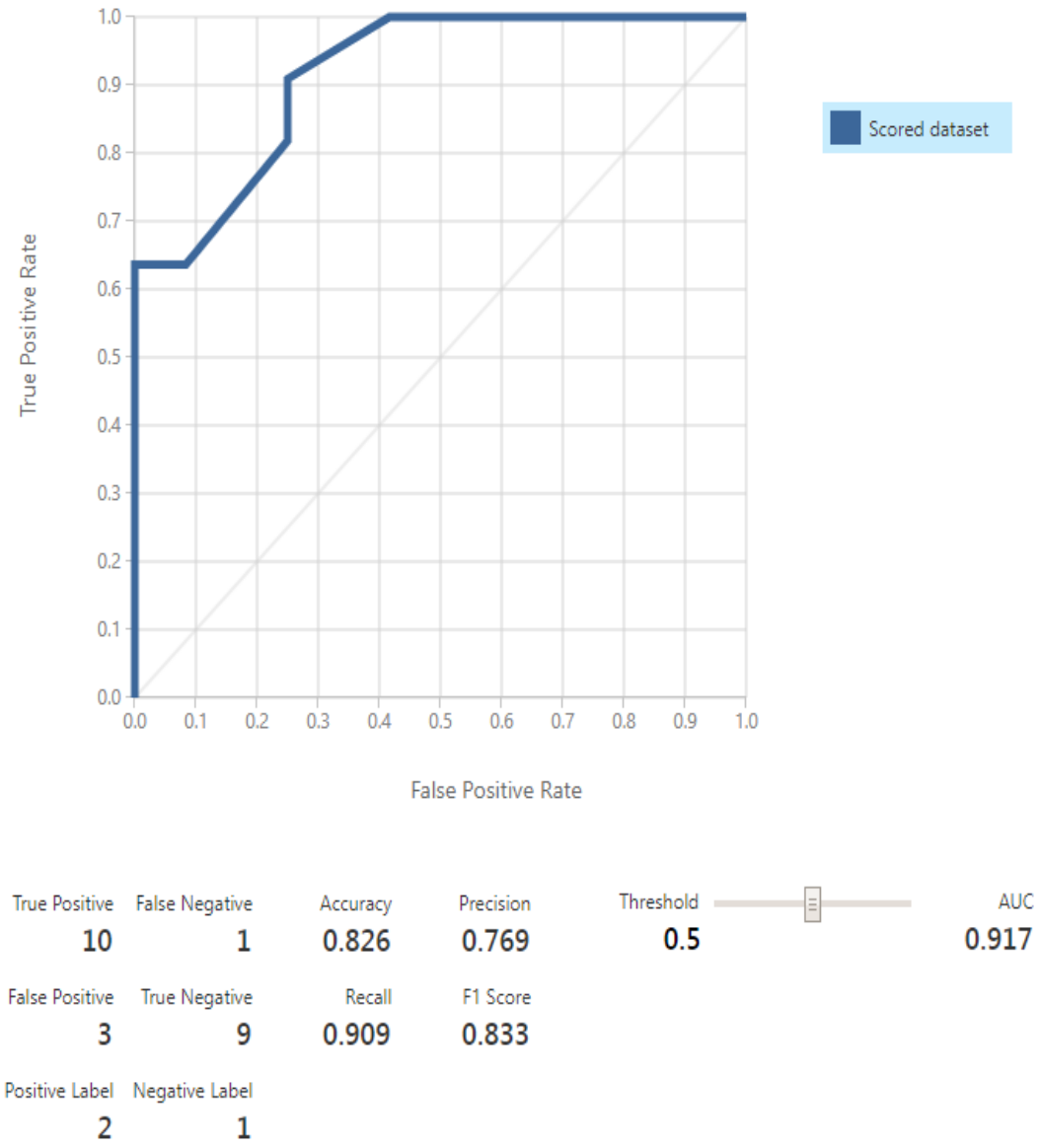


Figure 3.1.2: Predictive model evaluation

The below mentioned figure 3.1.1 shows how the results extract in the session. The question result shows separate to identify whether the patient has symptoms with the related disease. The patient results automatically saved in the database and it will automatically update when the doctor gets snapshot once again with the patient, but the question result does not update whether any modification with the question results.

Figure 3.1.3 : Showing the final results

3.2 Research Findings and Discussion

The main obstacle in completing this research was the collection of data samples. During this research there were certain constraints that can be faced as mainly it is the lack of dataset availability that would limit the machine learning algorithm from producing a more accurate and precise predictive model. As the dataset availability is low, the necessary had to be manually collected by visiting hospitals and making sure the appropriate patients are selected and the device developed for the project is used to collect the data samples that is used to train the model or else the accuracy level would be different if the samples are taken using some other resources which could alter the accuracy of the final results. Therefore, the building of the dataset by collecting data from scratch was an important aspect of the research. The proper organization of the data collected is also an important factor to note during the training of the predictive model. The collected data must be cleaned and transformed into the comma separated values format with each column labelled properly. During the data collection, the information collected should also be kept safe as the data are sensitive data and that information should be kept confidentially. If the server is hosted in the cloud, then a cost factor would be attached and in order to reduce it the services can be implemented locally in a server instance.

It was a challenge to configure the build the device in such a compact manner while

maintaining the consistency of images taken so as to preserve the performance of the system. With this compactness the overheating of the device became an issue as well. Therefore, measure including adding heat sinks to the device were taken to minimize the effect of heat on the operation of the device.

Another challenge faced was selecting the appropriate machine learning algorithm among the many algorithms and techniques which exist to classify data as needed in this research. Although there exists algorithms such as Artificial Neural Networks which can predict with extreme accuracy, those models require datasets with thousands of records. With the dataset for this research being limited to around 100 records, this posed a challenge in selecting a suitable algorithm which will produce results with acceptable accuracy. Thorough research and testing was made to finally settle on the decision jungle algorithm used in this research. This algorithm proved to be capable of producing acceptably accurate results with the limited dataset that was present.

It was also found that using a dedicated python server on a local machine was more efficient and much faster than using some other service such as firebase functions to connect the datastore with the web service of the machine learning predictive model. If using firebase functions, the time taken to gather the data from the datastore and to retrieve the result from the machine learning web service was significantly greater than using a custom python server to perform the same duties. This also meant that resources could be allocated to the python server as needed without having to consult third party providers.

Another challenge was finding the right balance between the amount of questions in the questionnaire while maintaining that each question is vital for accurate predictions. It was important that only the minimum number of most important questions were present while maintaining a high accuracy rate.

4. Conclusion

In conclusion the system produced by this research is able to predict the anemia condition of a user with acceptable accuracy. The dataset created by combining the image processing result and the answers to the questionnaire was used to yield a predictive model through machine learning which could predict the anemic status of a user. This system is up to a certain level, suitable for pre diagnosis of anemia without the need of blood testing. There is great difficulty in creating dataset for this type of research and the produced dataset in somewhat limited in size. With more time and effort , a more extensive and better dataset could be produced which would yield better accuracy in the predictions. Also, with a more extensive dataset, algorithms such as Artificial Neural Networks can be looked at to produce an even better predictive model.

The combination of raw objective data taken from a patient (image processing result) and subjective data from the same patient (questionnaire answers) can be used to enhance analysis done otherwise using only more limited inputs. Although this research is focused on the detection of anemia using this concept, it can be used as a template to be used in other research and projects to enhance their results.

In the end, the research objectives for this component as well as the overall research were achieved with reasonable results. There is optimism that this research and the dataset produced would inspire more research into the field of health informatics and thus help in improving lives all over the world.

5.References

[1] J. C. C.-F. R.-V. L. M.-C. Jaime Punter-Villagrasa, "Toward an Anemia Early Detection Device Based on 50-µL Whole Blood Sample," IEEE, Madrid, 2015.

[2] E. A. I. E. S. S. C. S. B. T. a. M. W. R. Bhattacharyya, "Towards low-cost, wireless blood anomaly sensing:" IEEE, San Diego, 2015.

[3] A. B. P. Sachin D. Khirade, "Plant Disease Detection Using Image Processing," IEEE, Pune, 2015.

[4] M. P. Meherwar Fatima, "Survey of Machine Learning Algorithms for Disease Diagnostic," Scientific Research Publishing Inc, Multan, Pakistan , 2017.

[5] webmd, "Understanding Anemia -- Symptoms," webmd.com, 01 January 2019. [Online]. Available: https://www.webmd.com/a-to-z-guides/understanding-anemia-symptoms#1. [Accessed 28 Feburary 2019].

[6] D. T. Jr., "Are you anemic," Everyday Health, 01 september 2014. [Online]. Available: https://www.everydayhealth.com/anemia/anemia-basics.aspx. [Accessed 12 02 2019].

[7] J. C. a. R. Nall, "Iron Deficiency Anemia," Healthline , 17 July 2017. [Online]. Available: https://www.healthline.com/health/iron-deficiency-anemia. [Accessed 13 Feburary 2019].

[8] W. H. Organization, "Global anaemia prevalence and number of individuals affected," WHO, 21 December 2008. [Online]. Available: https://www.who.int/vmnis/anaemia/prevalence/summary/anaemia_data_status_t2/en/. [Accessed 26 Feburary 2019].

[9] D. o. C. &. Statistics, "Prevalence of Anaemia Among Children and Women in Sri Lanka," Ministry of Healthcare & Nutrition, Colombo, Sri Lanka, 2006/07.

[10] S. Saxena, "Parallel Image Processing Techniques, Benefits and Limitations," researchgate, 5 January 2016. [Online]. Available: https://www.researchgate.net/figure/Basic-flow-diagram-of-the-steps-of-image-processing-Braeunl-2001_fig1_297629054. [Accessed 05 March 2019].

[11] S. S. S. A. Pooja Kamavisdar, "A Survey on Image Classification Approaches and," International Journal of Advanced Research in Computer and Communication Engineering, Bhilai, India, 2013.

[12] R. v. Loon, "Machine learning explained: Understanding supervised, unsupervised, and reinforcement learning," bigdata-madesimple, 05 Feburary 2018. [Online]. Available: https://bigdata-madesimple.com/machine-learning-explained-understanding-supervised-unsupervised-and-reinforcement-learning/. [Accessed 05 March 2019].