Journal of Engineering Science and Technology
Special Issue on SIET2022, May (2022) 58 - 76

AI-POWERED LOW-COST WEARABLE HEALTH TRACKER TARGETED TOWARDS ELDERLY IN DEVELOPING COUNTRIES

CELINE TIOH SU YIN, R. DHAKSHYANI*

School of Engineering, Asia Pacific University of Technology and Innovation (APU), Kuala Lumpur, Malaysia
*Corresponding author: dhakshyani@staffemail.apu.edu.my

Abstract

In recent years, more Malaysians are suffering from heart disease that can lead to major health consequences if symptoms are not detected at an earlier stage. As such, Artificial Intelligence (AI) is pivotal in detecting early signs of heart disease in the elderly through the various health data collected through Internet of Things (IoT). This paper proposes an IoT solution that is incorporated with a predictive system to prevent heart disease in elderly. This solution involves the use of an Arduino Uno and Photoplethysmography sensor, MAX30102, which measures heart rate and blood oxygen levels and sends them to a mobile application developed on the Node-Red platform for health monitoring and data collection. All parameters are collected via the application and sent to the AI model for analysis. The methodology of this system is centred around the heart disease prediction that is made from the data collected by the wearable device and its application. In the event of the presence of heart disease or the dropping of blood oxygen levels below a healthy threshold, the system makes record of it for medical analysis. This system aims at making primary medical attention available for all and providing lower income families and individuals with a more comprehensive yet affordable healthcare device.

Keywords: Artificial intelligence, Early heart disease prediction, General healthcare, Internet of things, Wearable health tracker.

1. Introduction

In a developing country like Malaysia where the economy is worsening and the cost of living is skyrocketing, health care is not as accessible and affordable as the ones provided in developed countries. Thus, many choose to neglect this aspect of life for spending on other necessities.

Heart disease is a major disease worldwide and the average age at which one suffers from heart disease has dropped to 58.6 years recently [1]. With recent advances in mobile sensor, wearable health tracker usually in the form of smart watches or fitness trackers have been more widely accepted as an addition to the existing health data collection methods in medical research [2]. The combination of hardware and software components together with an AI-powered predictive system provides a cheaper alternative than the smart watches and fitness trackers in the market to meet the need for affordable and comprehensive healthcare solutions in poorer countries.

The main sensor chosen is the Photoplethysmography sensor which makes up an inexpensive, non-invasive optical instrument. It detects both heart rate and blood oxygen in a single device. The AI predictive system acts as an early heart disease prediction system which predicts the possibility of the presence of heart disease in a person. The two components in this project is integrated with a mobile application which displays all sorts of health data and further processes them before passing the data as in Fig.1.

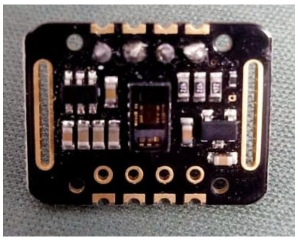

Fig. 1. Photoplethysmography sensor (MAX30102).

The Photoplethysmography sensor is often used to obtain and monitor heart rate and blood oxygen levels. It measures the changes in blood volume by detecting the changes in light intensity passing through the surface of skin. The device comprises of a light source and a photodetector. The light source for the chosen sensor includes the infrared (IR) and the red LEDs, while the photodetector is a phototransistor or a photodiode, LDR as in Fig. 2.

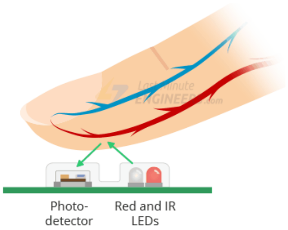

Fig. 2. Components in the Photoplethysmography sensor [3].

2. Research Problem

According to the statistics by news outlet CNBC, the older demographic is not making use of healthcare technologies such as fitness trackers and smart watches [4]. Moreover, though smart watches are priced a little bit higher, both fitness trackers and smart watches are usually overpriced and lacking in healthcare functions. The cheaper range of fitness trackers, on the other hand, are severely lacking in useful functions for healthcare for elderly, usually without heart rate monitor and GPS [5]. However, as the common health problems present in elderly are hypertension, heart disease, and diabetes [6], existing fitness trackers and smart watches will not serve a purpose for the older demographic in poorer countries as they are. This is because they cannot fulfil the criteria of affordability and functionality. Most devices are overpriced, and the cheaper ones do not come with useful functions for elderly such as detecting irregularities in heart rate and blood oxygen. Furthermore, collection of private data for medical treatment and diagnostics have not been explored much as of yet. Lastly, sole reliance on humans alone can lead to human errors such as overlooking of underlying health conditions which can have major negative impacts on a person and their quality of life in general.

3. Related Work

Nashif et al. [7] discusses and does a comparison of various machine learning algorithms for a heart disease prediction system using both the Cleveland and the Statlog Heart Disease Dataset as dataset for training. The algorithms of discussion include Naïve Bayes, Artificial Neural Networks (ANN), Support Vector Machine (SVM), Random Forest, and Simple Logistic Regression.

Naïve Bayes is a statistical classifier which assumes that there are no dependency between characteristics of data and is a powerful algorithm used in predictive modelling. The assumption of no dependency between data characteristics improves data classification through maximization of posterior probability. As such, this algorithm is, in theory, able to achieve the lowest error rate when compared with other algorithms. However, the assumption of no class dependency and the lacking in available probability data may prove otherwise practically.

ANN or multilayer Perceptron is an algorithm that is capable of making complex decisions and is a major tool in machine learning. It comprises of three layers: input, output, and hidden layers. Next, the Simple Logistic Regression is an algorithm which predicts from a field of statistics and is mainly used for binary classifications or two-class predictions. Logistic regression works best where a relationship between attributes and output is not necessary.

The Random Forest also known as Bagging or Bootstrap Aggregation is one of the most renowned algorithms for learning as it is a potent statistical method to prediction of a value from a sample of data. It predicts by averaging out large number of samples to increase the accuracy of each prediction. Although, Bagging or the Bagging Trees method is similar to bootstrap, it implements multiple decision trees instead of estimating the mean of every data.

Lastly, Support Vector Machine (SVM) is a method which uses both linear and non-linear data. Nonlinearity is used through non-linear mapping method to turn training data to a higher dimension of data. Among the machine learning algorithm evaluated by Nashif et al. [7], the SVM algorithm showed the highest accuracy of 97.53%, followed by Random Forest, Simple Logistic Regression, Naïve Bayes, and ANN with recorded accuracies of 95.76%, 95.05%, 86.40% and 77.39% respectively [7].

4. Investigation of Tools and Techniques

4.1. Photoplethysmography sensor

Table 1 shows the comparison of the specifications of the Photoplethysmography sensors taken into consideration. The MAX30102 and BH1792GLC sensors appear to be more suited for this application due to their high sensitivity to changes in light intensity which gives more accurate results, lower power consumption for a longer lasting battery life, high sampling rate for real-time detection, and low-noise capabilities to remove unwanted ambient light.

Table 1. Specifications of Photoplethysmography sensor.

Sensor	MAX30102	SEN0203	BH1792GLC
Output readings	High sensitivity	Less accurate and consistent readings	High sensitivity
Connectivity	I2C compatible interface communication	Gravity interface for easy plug-and-play connectivity	I2C compatible interface communication
Power consumption	Low power consumption (<1mW)	Higher power consumption (~0.054W)	Low power consumption
Sampling rate	High sampling rate	Moderate sampling rate	High-speed frequency sampling (1024Hz)
Susceptible to noise	Low-noise with ambient light rejection	Has minimal noise	Low-noise due to optical filter

Although the SEN0203 sensor appears to be lagging behind in terms of performance as compared to its counterparts, it is still sufficiently suitable in terms of power consumption, sampling rate and susceptibility to noise [8]. In view of this discussion, the sensor chosen is the MAX30102 sensor. Although similar in performance to that of the BH1972GLC sensor, the price of the MAX30102 is the deciding factor in this case for the construction of a low-cost device to target consumers of poorer countries.

4.2. Programming language

The performance and availability of both programming languages are both almost identical for this case. However, MATLAB stands out from the Python language due to its wide selection of libraries and built-in functions which can be downloaded with simple steps and for free with the academic license compared to Python. This is due to the reason that MATLAB is a software made for academic learning purposes. Furthermore, building an AI model is very simple with the Classification Learner App provided within the software. Table 2 shows the comparison of the programming languages.

Table 2. Comparison of programming languages for building AI algorithm.

Language	Python	MATLAB
Code difficulty	Up to personal preferences	
Libraries for machine learning	TensorFlow, Keras, PyTorch	Built-in with Statistics and Machine Learning Toolbox
Data visualisation and mathematical functions	NumPy, SciPy, Scikit-learn, Pandas	Built-in
Availability	Open Source	Proprietary (Academic license provided)

4.3. IoT platform for mobile application development

The three IoT platforms were taken into consideration for their simplicity to develop for a beginner at mobile application development. Although all three platforms are relatively easy to code to develop mobile application and are supported on Android, the main consideration that differentiates the three platforms is the GUI design appearance and flexibility followed by cost. From Table 3, the platform which provides the best designs for GUI in terms of appearance and flexibility for higher level of customisations is Node-Red. Moreover, it is free to use as opposed to GraspIO which requires the use of very specific microprocessor, the Raspberry Pi, and a GraspIO Cloudio extension which adds cost to the overall project.

Table Error! No text of specified style in document.. **Comparison of IoT platform for building a mobile application.**

IoT platform	MIT App Inventor	GraspIO	Node-Red
Programming language	Block programming	Drag-and-drop programming	Flow-based programming
GUI capabilities	• Limited components for UI design • Supports only outdated and old-fashioned designs	• Less rigid designs but limited to functions provided within the platform • Not flexible for customisations	• Advanced designs yet simpler than traditional mobile application development platforms to configure
Mobile app support	MIT AI2 Companion	GraspIO Studio	Termux

5. Proposed Methodology

5.1. Photoplethysmography sensor (MAX30102) data collection

Figure 3 represents the flow chart of the operational steps for the hardware part of this project which consists mainly of the operations of the MAX30102 sensor whereby the user's heart rate and blood oxygen levels are measured. If finger is applied to the sensor, it prints the data every time the heart rate or blood oxygen levels change, else it prints a string 'null' once.

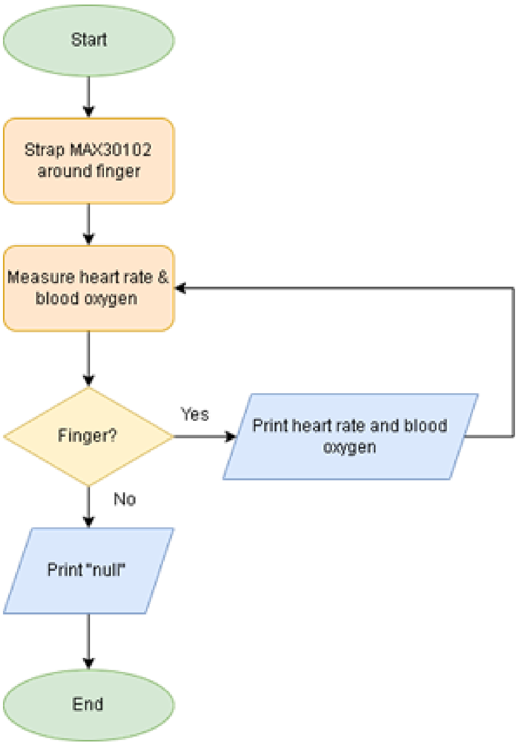

Fig. Error! No text of specified style in document.. **Flow chart of photoplethysmography sensor.**

5.2. Mobile application

The mobile application in Fig. 4 which holds the entire system together has the most complex operational flow with multiple conditions and functions. When a user is trying to log in to the application, it first checks if an account for that user exists. If it does not exist, the user would have to sign up for an account first by entering some of the credentials as mentioned in "Mobile application development" under the 'signup' page before successfully logging in. However, if the account exists, the user would be successfully logged into the application. Within the application, the user is allowed to freely navigate between the three main pages, namely 'dashboard', 'schedule', and 'profile'. The 'dashboard' page allows users to view daily schedule, heart rate and blood oxygen level as well as log out of the application as desired. The 'schedule' page allows users to view the schedules of every day and add items to the schedule where the user is allowed to edit or delete an item if that item exists. If it does not exist, the user will be required to add the item first. In the 'profile' page, however, it simply displays the user information given by the registered user at the 'signup' page and allows user to update their profile by providing their blood pressure and blood sugar readings.

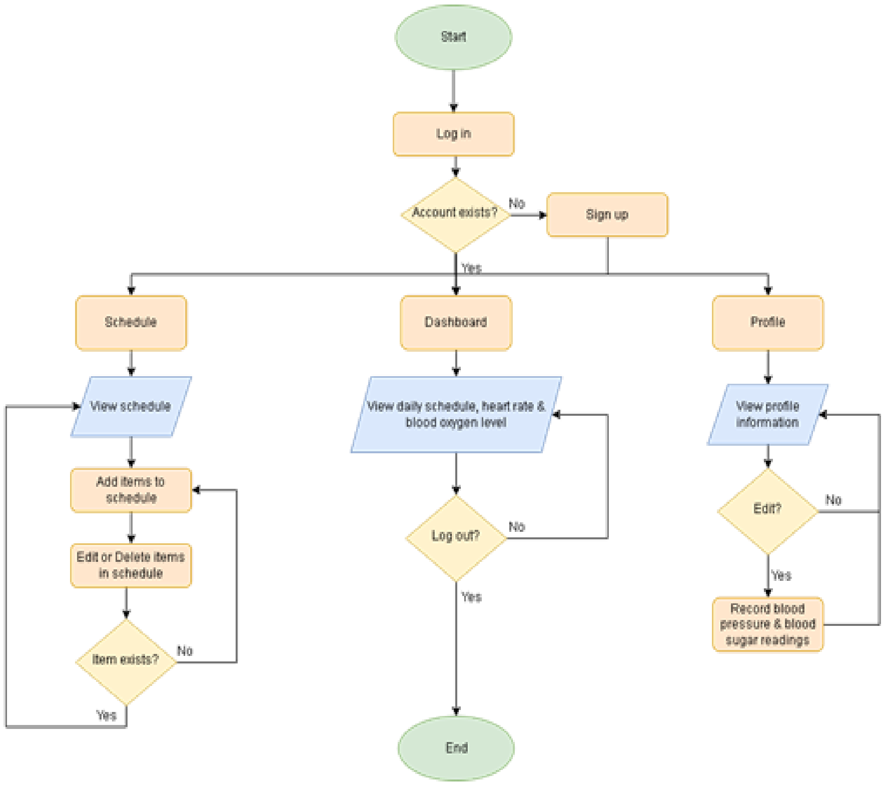

Fig. 4. Flow chart of mobile application.

5.3. Heart disease prediction model

Figure 5 shows that after collecting the users' data, medical personnel can assess the collected data by running it through MATLAB whereby the program will automatically load the files into the workspace and prepare the dataset by filtering out unwanted data before passing it into the model for prediction. The predicted value 1 indicates the presence of heart disease and 0 indicates the absence of heart disease. If the number of predicted values has more 1 than 0, the model will classify the user as having heart disease and vice versa. Notwithstanding, the accuracy of the model on this set of data is also calculated and displayed to ensure integrity of data.

Before collecting data from participants, the participants are briefed about the procedures which would be required from them to follow as well as the criteria for their participation as stated in the consent form. The sensor is then strapped to the participant's finger with a Velcro tape so that constant pressure and position can be applied for more accurate recording of data. This is because blood flow in finger is different with different pressure applied by the finger on the sensor and different positions of the finger affects how the infrared and red light is shone onto the surface of skin, thus affecting the reflection of light captured shown in Fig. 7.

The entire session is recorded over a duration of 5 minutes excluding the time taken to brief the participants. The data is sent and saved to the mobile application at Node-Red concurrently as the sensor is being strapped around the finger of the participant.

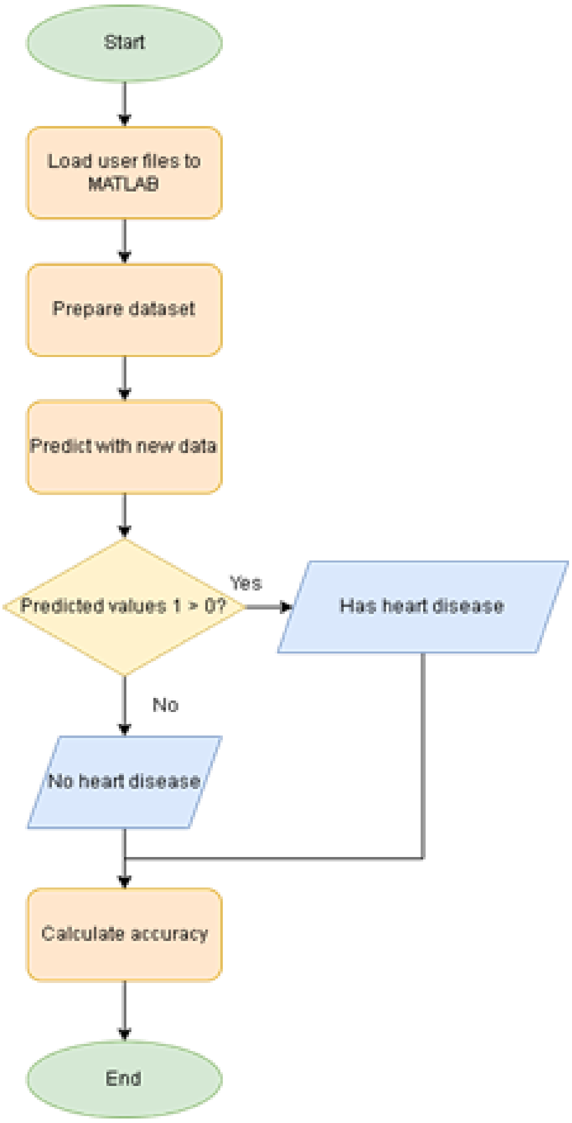

Fig. 5. Flow chart of heart disease prediction model.

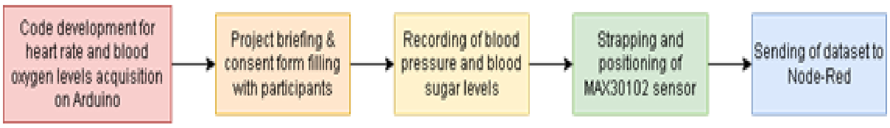

Fig. 6. Block diagram of sensor testing and dataset acquisition.

Fig. 7. Block diagram of classification mode development.

5.3.1. Classification model development

The other main component of this project is the classification model which is developed on MATLAB and tested for accuracy using the Classification Learner App in MATLAB. This allows the training and testing of data with several if not all types of models available.

This stage in implementation first starts with selection of useful and significant features from the training dataset. The features selected are age, sex, resting blood pressure, fasting blood sugar, and heart rate. This is because some of the selected parameters are more significant in the prediction of heart disease. The selection of parameters is also dependent on how feasible and practical it is to collect those data. Moreover, taking into consideration redundant or irrelevant attributes from the dataset will also affect the accuracy of model training due to dimensionality issues where algorithms struggle to train effective models due to the large number of insignificant attributes compared to the number of observations.

To further improve the consistency of accuracy of model between training and testing set, cross-validation is used to prevent overfitting problem which was faced in the initial stages. The k-number of folds is decided to be 5 because it showed the highest performance in the model training. This partitions the dataset into k subsets to generate multiple mini train-test data where hyperparameters are tuned according to the splits.

The training of model is tested on all available models in the Classification Learner app in MATLAB. The three models that showed the highest accuracy on the dataset selected are the Ensemble-Bagged Trees, Wide Neural Network, Weighted KNN with accuracies of 85.3%, 85.0%, and 84.5% respectively. Although for the testing set, the Medium Neural Network model showed highest accuracy which is 90.1%, Ensemble-Bagged Trees model is not far behind, placed at second highest which is 89.9%. The other two models which showed highest accuracy among models on training set did not do as well on the testing set as the Bagged Trees model with accuracies 87.7% for both. Hence, the Bagged Trees model is chosen as it still showed best overall accuracy of 87.6% on both training and testing set.

After the Bagged Trees model is selected, the dataset from the sensor and mobile application is prepared by removing unwanted attributes, filtering out unusable data, and various data conversions. The unwanted attribute to be removed is password due to privacy concerns. Notwithstanding, it is also not required for heart disease prediction. Furthermore, unusable heart rate data is filtered out to avoid having them affect the accuracy of prediction. This unusable data comes from the instability of data due to inconsistent pressure and position as mentioned in the first part of this chapter in “Sensor testing and dataset acquisition” as well as a bug in the printing data sent from HC-05 module to serial of laptop. Various data conversions are also performed for converting data that is collected to the format that is accepted by the model according to the format provided by the training dataset. These conversions include converting the attributes ‘sex’, ‘fbs’, and ‘num’ from mobile application to a binary number to pass into the model for prediction. The ‘sex’ attribute is initially provided by the mobile application as male or female strings inputted by the participants into the app, while the 'num' attribute is passed as Boolean values from the mobile application. The 'sex' data accepted by the prediction model is 1 for 'male' and 0 for 'female', whereas 'num' is categorised by 1 for presence of heart disease and 0 for absence of heart disease. The 'fbs' attribute was initially numeric values of the participants' blood sugar levels which was then converted to 1 for values above 120, else 0.

The evaluation of prediction model on how well it performed on new data in terms of accuracy is calculated within the system by finding out the confusion matrix.

5.3.2. Mobile application development

The mobile application development process is divided into four parts with each process representing a function or a page in the mobile application shown in Fig. 8. The login and signup page are similar and relatively easier to create. As such, they were grouped as one. When a user first opens the mobile application, they will find the 'login' page which asks for their login credentials. If no login credential is provided, user can click on the 'signup' button to be directed to the 'signup' page where they can sign up for the mobile application by providing basic information such as their username, password, age, and sex. The 'login' page grants access to the mobile application and collects the users' name which allows the mobile application to display, collect data as well as perform certain functions which are specific to the user such as displaying the user's schedule for the day, collecting their heart rate, and saving it to a CSV file attached to their username and so on. In the 'signup' page is where the user's information is created and added to the global data 'accounts'. This global data is responsible for the saving of all users' information to a database where data can be retrieved and used. Moreover, empty arrays for the newly registered users are also added to the global data 'schedules' for the application to be able to push new schedule data into the global data in the schedules page. This allows users to view schedules specific to the user that is logged in.

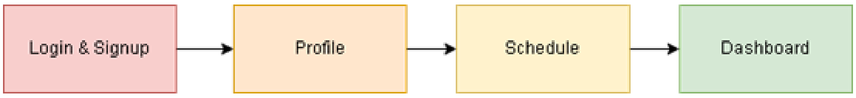

Fig. 8. Block diagram of mobile application development.

The page in the mobile application that is developed next is the 'profile' page which displays the information added by users at the 'signup' page as well as allows them to update their profile with blood pressure and blood sugar data. Within this function, the heart rate passed into the mobile application from the sensor is also pushed into the 'accounts' global data for saving the file to a filename generated by the logged in username. If the filename for that particular exists, a file lister will write the data into the existing file, else a new file with a new filename is created with column headers declared to be 'username', 'password', 'age', 'sex', 'trestbps', 'fbs', 'thalach', and 'num'.

The 'schedule' page in the mobile application is rather the most complicated page compared to the other pages. When the 'schedule' page is selected, the user will reach the display page which displays the schedule of the selected day. Within that page also exists the three buttons to modify the contents of the schedule, and they are to add, edit, and delete the items in the schedule. This function is where the empty arrays in the 'schedules' global data created for the particular user in the 'signup' function as mentioned in the first paragraph is pushed into with items that the user so desires. The 'add' function allows users to add medicine and its time to the schedule, the 'edit' and 'delete' functions include a dropdown which finds the items in the schedule for the chosen day where users can edit the time and date of the item or just delete them altogether.

Lastly, in the 'dashboard' page, users can find the schedule displayed for the day which the user logged into. Furthermore, the heart rate received from Arduino is displayed as a time series line chart and the blood oxygen level as a numeric value where the numeric value will give a warning and appear in red if the parameter reaches below a certain threshold. The schedule on the dashboard is activated by two ways, the 'login' or 'signup' buttons or the 'dashboard' button. The 'dashboard' page is also the only page which gives the option to log out of the mobile application where a confirmation dialog is prompted to ensure the user is certain of their actions.

5.3.3. Integration of components

As shown in Fig. 9, the process of integrating the components involves connecting the main components of this project and fine tuning of each component to ensure the integration is cohesive. The first integration process involves connecting the hardware component to the mobile application in Node-Red through establishing the connection of the HC-05 Bluetooth module to the serial port of the laptop which acts as the server.

Fig. 9. Block diagram of integration of components.

However, with the Bluetooth connection established, a few issues were encountered. As such, fine tuning is made to the Arduino codes and the fetching of data within Node-Red. The Arduino codes was modified to only print data when there is a change in value so to improve in the display of data in Node-Red. Not only that, but the data also coming into Node-Red is required to be split into arrays or into two separate entities, heart rate and blood oxygen so that they can be used to perform separate functions.

While the blood oxygen value is used within Node-Red alone as a precaution to the user, the heart rate is saved together with the other data collected in profile as dataset for heart disease prediction. This is done by compiling the profile information collected with the heart rate sent from Arduino into CSV files attached to a particular user.

The CSV files created for the users are loaded to the MATLAB workspace for medical personnel to perform health assessments and heart disease predictions at any time with the data collected. Notwithstanding, the new predicted values are appended to the participants' files as a copy for evaluation of the participants' health data. Since data is collected and saved in real-time, medical personnel can access the files with up-to-date data at any time and perform new predictions where necessary.

6. Experimental Setup

6.1. License plate recognition system setup concept drawing

The recording session for dataset collection requires setting up of hardware such as computer, MAX30102, supplying power from power bank and establishing HC-05 connection. Fig. 10 shows the recording session for the subject’s index finger of left hand.

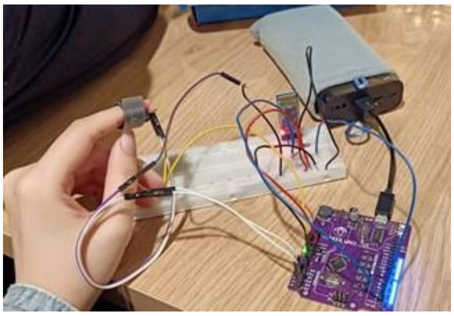

Fig. 10. Setup of hardware during dataset recording session.

7. Simulation Results

Tables 4 to 6 shows the repeatability results.

7.1. Repeatability

Repeatability test performed in this project is only assessed on the 65 samples extracted in random from each subject and not the entire data. This is because the total amount of data for each subject is not the same, hence difficult to compare.

Table 4. Prerequisites of constant for repeatability test on MAX30102 and heart disease prediction model

Index	Condition	MAX30102	Heart disease prediction model
1	Measurement function to test	Heart rate	Presence of heart disease
2	Measurement range	20-120	0 & 1
3	Test point(s)	10% & 90%	10% & 90%
4	Method	Strap on MAX30102, attach pulse oximeter, connect to mobile application and Bluetooth	Select data to assess, predict and analyse
5	Equipment	Photoplethysmography sensor (MAX30102)	MATLAB
6	Operator	Researcher	Researcher

Table Error! No text of specified style in document.**. Repeatability results for MAX30102.**

Subject	Mean	Standard deviation	Degrees of Freedom (DOF)	Weighted variance (SS)
1	78.56923	4.821726	64	1487.938
2	88.70769	7.33298	64	3441.446
3	50.61538	12.31195	64	9701.385
4	74.23077	6.896034	64	3043.538
5	70.03077	10.67264	64	7289.938
Total			320	24964.25
Pooled standard deviation (Sp)			**8.832512**	

Table 6. Repeatability results for heart disease prediction model

Subject	Mean	Standard deviation	Degrees of Freedom (DOF)	Weighted variance (SS)
1	0	0	64	0
2	0.015385	0.124035	64	0.984615
3	0	0	64	0
4	0	0	64	0
5	0	0	64	0
Total			320	0.984615
Pooled standard deviation (Sp)			**0.05547**	

7.2. Accuracy

Tables 7 and 8 shows the accuracy results while Table 9 shows specificity results.

Table 7. Accuracy results for MAX30102.

Subject	Accuracy (%)
1	70.83
2	62.11
3	19.51
4	54.17
5	95.76
Average	**60.48**

Table 8. Accuracy results for heart disease prediction model.

Subject	Accuracy (%)
1	71.7
2	98.25
3	100
4	100
5	100
Average	**93.99**

7.3. Specificity

Table 9. Specificity results for heart disease prediction model.

Subject	Specificity (%)
1	71.7
2	98.25
3	100
4	100
5	100
Average	**93.99**

7.4. Reliability

Table 10 shows the level of agreement and Table 11 shows reliability results. The heart disease prediction model works best with subjects 3, 4, and 5 which are all young adult males. The performance of model on female subjects 1 and 2 are less excellent, more so with subject 1 shown in Fig. 11. This is because subject 1 has different blood pressure readings in their dataset which indicates the weight or the role of blood pressure readings in the prediction of heart disease aside from heart rate readings.

Table 10. Level of agreement based on Kappa score.

Kappa score	Level of agreement
<0.20	Poor
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Good
0.81 – 1.00	Excellent

Table 11. Reliability results for heart disease prediction model.

Subject	Kappa score	Level of agreement
1	0.43	Moderate
2	0.97	Excellent
3	1.0	Excellent
4	1.0	Excellent
5	1.0	Excellent
Average	**0.88**	**Excellent**

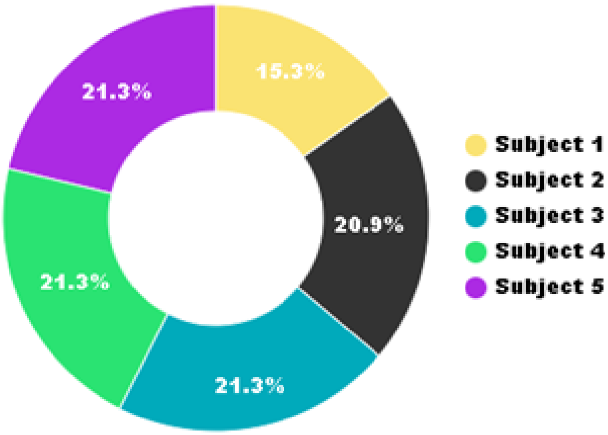

Fig. 11. Performance on model's accuracy on each subject.

8. Data Analysis

As observed from the MAX30102 sensor accuracy tabulated in Table 12, the accuracy is significantly lower when the number of samples is too few. This is because there are too few samples to compare with the actual values from the clinically tested Yobekan pulse oximeter. As such, a solution to this would be to increase the number of samples for each user to at least 200 samples for better accuracy.

Moreover, Tables 13 to 15 shows that the attributes 'sex' and 'blood pressure' have significant impacts on the prediction results. The effect of blood pressure was observed in the results for subject 1 which showed variations in blood pressure with the rest of the parameters almost consistent. These two scenarios were caused by the dataset used to train the model which consisted mainly of males and subjects with blood pressures of 100 mmHg and above. This resulted in the weakness of the model to predict some datasets accurately.

Table 12. Accuracy for MAX30102 sensor affected by number of samples.

Component	Accuracy (%)				
	848	171	65	144	283
MAX30102 sensor	70.83	62.11	19.51	54.17	95.76

Table 13. Average accuracy for heart disease prediction model affected by sex.

Component	Average accuracy (%)	
	Male	**Female**
Heart disease prediction model	100	84.98

Table 14. Prediction for heart disease prediction model affected by blood pressure.

Component	Presence of heart disease		
	99 mmHg	**110 mmHg**	**120 mmHg**
Heart disease prediction model	Yes	No	No

Table 15. Prediction for heart disease prediction model affected by heart rate.

Component	Presence of heart disease	
	<110 bpm	>110 bpm
Heart disease prediction model	No	Yes

Out of the 484 subjects recorded in both the Cleveland and Statlog heart disease dataset which were used to train the model, 333 of them are male while only 151 of them are female subjects. As a result, this unequal number of male and female representation has caused the model to be weaker when predicting results for female subjects as shown in Tables 16 and 17.

By comparing the confusion matrix of male and female subjects, it can be concluded that the model has no problem identifying the diagnosis of heart disease for male subjects. However, it falsely classified 243 instances of the predicted value as having heart disease.

Table 16. Confusion matrix for male subjects.

		Expected	
		Has heart disease	**No heart disease**
Predicted	Has heart disease	0	0
	No heart disease	0	462

Table 17. Confusion matrix for female subjects.

		Expected	
		Has heart disease	**No heart disease**
Predicted	Has heart disease	0	243
	No heart disease	0	776

Notwithstanding, it also failed to classify certain data when the blood pressure readings drop below a certain threshold. This is because in the Cleveland and Statlog heart disease dataset, there are only 3 instances where the blood pressure readings drop below 100 mmHg which is at 94 mmHg. As such, it is not as skilful in predicting subjects with blood pressure below a 100 mmHg.

As subject 1 is the only subject with variations in blood pressure and blood pressure below 100 mmHg, the confusion matrix of the model’s performance on this subject is observed in Table 18. Subject 1 tested 3 different blood pressure readings, namely 99, 110, and 120 mmHg. From Table 18, 240 instances of false positives in the confusion matrix all belong to the blood pressure of 99 mmHg, while the other two recorded blood pressure readings gave true negative values.

Table 18. Confusion matrix for subject 1.

		Expected	
		Has heart disease	No heart disease
Predicted	Has heart disease	0	240
	No heart disease	0	608

With regards to this issue, a possible solution would be to implement reinforcement learning to the current model to retrain the model to be able to predict with a variety of parameters.

9. Comparison to Previous Research

The contribution in this project as compared to other similar projects is the inclusion of a mobile application which collects data and sends them to the AI model for predictions in real-time. Most research only focus on creating the AI model while some of them collect data through a website. With the use of this application paired with the sensors from the device, users can monitor their conditions comfortably from their own home or wherever they are which provides the elderly and the ill the quality of life that is usually assumed to be lost with the presence of an illness. Not only that, it provides responsible individuals such as the patients' children or relatives with an ease of mind being able to monitor their loves ones personally.

The outcome of this project also yielded best results with accuracies of above 80% in both training and testing set as well as accuracies of as high as a 100% on certain subjects. Notwithstanding, this project also contributes to real-time analysis of various health parameters such as heart rate and blood oxygen which are important parameters to observe especially in elderly people. Besides, it also provides responsible individuals or medical personnel with an extensive report of the early predictions of heart disease in the users which could prevent heart disease from worsening or early diagnosis of heart disease by recognising trends.

Lastly, the proposed system is also constructed to be extremely affordable compared to the fitness trackers and smart watches in the market which do not suit medical applications at all. By focusing more on the functions of the system and not the aesthetics or the visuals, useful components that are non-invasive and inexpensive were integrated with a mobile application that reduces the burden of those responsible of monitoring the patients as well as an algorithm that does early diagnosis of the patient's heart condition.

10. Conclusion

The proposed system has achieved all the research objectives of this project. A prototype with heart rate and blood oxygen readings were constructed with affordability in mind. These readings were successfully sent to a mobile application developed with multiple uses such as preparing a schedule to take medicine easily, display health statistics and health warnings, and collect data for early heart disease prediction on classification model which predicts whether the user is suspected to be diagnosed with heart disease. The best results for accuracy for the Photoplethysmography sensor are 95.76% with proper positioning of the sensor and for the heart disease prediction model is 100% on healthy subjects.

11. Limitations

One of the limitations of this project is the instability of the MAX30102 sensor in detecting heart rate. This is because the instrument used is an optical instrument which gets affected by pressure applied by fingertip on sensor and ambient light. This resulted in a reduction in accuracy of heart rate data for some subjects. Another limitation in the MAX30102 sensor is the lack of samples for heart rate which results in lower accuracies and higher fluctuations in data.

A limitation in this project is also the lack of real-time data for blood pressure and blood sugar readings which causes the prediction to be less accurate as it gives readings that might be outdated. Furthermore, another limitation of this project is the accuracy of the heart disease prediction with certain parameters such as sex and blood pressure where the accuracy drops when the subject is a female and when blood pressure drops below 100 mmHg. This is because of the lack of variations in the dataset for those attributes for the model to train on. Hence, the model is weaker in classifying those data.

Lastly, another limitation is the lack of representation of heart disease patients in the dataset due to ethical issues which come with testing of system on patients. This results in the data having only true negatives for values that are predicted correctly.

12. Recommendation and Suggestions

Some suggestions for improvement on the limitations that appeared in the MAX30102 sensor is to increase the number of samples and allow the sensor readings to stabilise over time. A more secure clip or strap can also be used instead for more consistent pressure from fingertip on sensor.

Moreover, reinforcement learning can be implemented on this system to improve the accuracy of the prediction model. This is to retrain the model on new data so that it can predict with a wider variety of parameters and conditions to build a more robust classification model.

On the hardware side, GPS can also be included together with an emergency alert system to locate patients for emergency purposes. This would allow the users to contact emergency lines or loved ones easier as well alert responsible individuals of emergencies or abnormalities in the user's condition. A heart disease information page which contains information from credible sources can also be included to provide reliable information that can guide users in taking care of their health.

Acknowledgement

Firstly, I would like to express my deepest gratitude to my supervisor, Ir. Ts. Dr. Dhakshyani Ratnadurai, for the guidance, patience, support, advice, and constructive criticism that was offered to me during both the investigation and implementation phase of my final year project. Her constructive and gentle feedback and comments gave me courage and guidance to complete my project. Besides that, I am very grateful to my second marker, Assoc. Prof. Ts. Dr. Shankar Duraikannan for his forbearance and his feedback on my report writing and research skills. He also highlighted many concerns which gave me greater insight on my project.

I would also like to thank my friends from Asia Pacific University: Aiman Haziq, Teoh Kean Hao and Mak Yuin Onn for helping, moral support as well as help in participating in the testing of my system. Finally, I would like to thank my family for supporting me morally and financially throughout the years. They have also helped me in many ways by participating in the testing and giving feedback on the mobile application.

References

1. Ramakrishnan, N. (2019). More Malaysians are suffering heart attacks at a younger age. Retrieved October 16, 2019, from https://www.thestar.com.my/lifestyle/family/2019/10/16/heart-attacks-not-only-the-old.
2. Henriksen, A.; Mikalsen, M.H.; Woldaregay, A.Z.; Muzny, M.; Hartvigsen, G.; Hopstock, L.A.; and Grimsgaard, S. (2018). Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. *Journal of Medical Internet Research*, 20(3).
3. Last Minute Engineers. (2022). Interfacing max30102 pulse oximeter and heart rate sensor with arduino. Retrieved October 30, 2021, from https://lastminuteengineers.com/max30102-pulse-oximeter-heart-rate-sensor-arduino-tutorial/.
4. Gustafson, K. (2015). Smartwatch or fitness tracker? Why age, sex matter. CNBC. Retrieved October 30, 2021, from https://www.cnbc.com/2015/01/07/smartwatch-or-fitness-tracker-why-age-sex-matter.html.
5. Peckham, J.; and Ellis, C. (2022). The best cheap fitness trackers 2022: budget activity bands from Fitbit and more. techradar: Retrieved August 14, 2017, from https://www.techradar.com/sg/best/best-cheap-fitness-trackers.
6. Abbott nutrition. (2021). Top 5 health concerns of the elderly and how to resolve them. abbott nutrition. Retrieved July 20, 2018, from https://abbottnutrition.com.my/adult-nutrition/5-health-issues-elderly-how-to-resolve-them?gclid=Cj0KCQjwxtSSBhDYARIsAEn0thQJ-NC1HJm1TQu7ZAF2M1nHiwyWXYcczPInEyZKHeha7lN90lU4-SYaAm0wEALw_wcB&gclsrc=aw.ds.
7. Nashif, S.; Raihan, M.R.; Islam, M.R.; and Imam, M.H. (2018). Heart disease detection by using machine learning algorithms and a real-time cardiovascular health monitoring system. *World Journal of Engineering and Technology*, 6(4), 854-873.
8. Holko, M.; Litwin, T.R.; Munoz, F.; Theisz, K.I.; Salgin, L.; Jenks, N.P.; Holmes, B.W.; Watson-McGee, P.; Winford, E.; and Sharma, Y. (2022). Wearable fitness tracker use in federally qualified health center patients:strategies to improve the health of all of us using digital health devices. *Npj Digital Medicine*, 5(53), 1-6.