Journal of Advanced Research in Applied Mechanics 105, Issue 1 (2023) 41-57

SEMARAK ILMU PUBLISHING

Journal of Advanced Research in Applied Mechanics

Journal homepage:
https://semarakilmu.com.my/journals/index.php/appl_mech/index
ISSN: 2289-7895

HealthyHeart Data Visualization: Predicting Heart Condition Using Machine Learning

Masturina Nazurah[1], Zatul Amilah Shaffiei[1,*], Nor Aziah Daud[1], Nor Diana Ahmad[1,♣], Zatul Alwani Shaffiei[2]

[1] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
[2] iKohza Software Engineering for Industrial Revolution (SEiR), Malaysia - Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 15 January 2023
Received in revised form 25 March 2023
Accepted 3 April 2023
Available online 20 April 2023

Keywords:
Heart disease; predictive model; machine learning; data visualization; big data

ABSTRACT

This paper deals with the primary cause of mortality in society which is heart disease. Neglecting this disease will expose you to incredibly serious risks and consequences. The factors that contribute to the increment of mortality are lifestyle and lack of awareness in society. Heart disease can strike anyone, regardless of age or gender. What makes it different is the risk level of each factor. Due to that, an awareness campaign towards this disease must be taken seriously so that early diagnosis of this disease can be made to avoid harmful scenarios. Thus, HealthyHeart comes to the rescue, where the features included in this dashboard are predictive models using machine learning to predict heart conditions. To complete the dashboard, Agile methodology has been used with the OSEMN model which focuses on the public user needs. Finally, the finding shows that HealthyHeart is a good dashboard as it provides information about users' heart disease risk assessment and added information that can help them to get educated about heart disease and its risk. By knowing this knowledge, people will be aware of their health and take a precautious step to prevent and maybe reduce heart disease cases.

1. Introduction

The heart is one of the most important organs for a human to live besides the brain. It is an organ that distributes oxygenated blood throughout the human body which is very important for human life [1]. Several factors can lead to the failure of the heart such as smoking, obesity, physical inactivity, nutrition, high blood cholesterol, high blood pressure, diabetes mellitus, metabolic syndrome, kidney disease, and sleep disorder [2]. However, it can be maintained by living a healthy lifestyle and exercising regularly. Negligence of its health will bring disaster like death.

* *Corresponding author.*
E-mail address: zatul@fskm.uitm.edu.my
♣ *Corresponding author.*
E-mail address: nordiana@tmsk.uitm.edu.my

https://doi.org/10.37934/aram.105.1.4157

In addition, data stated from World Health Organization (WHO) highlighted that about 85% of deaths from 17.9 million lives were taken each year in the United States where it represents 32% of global deaths due to cardiovascular diseases (CVDs). Meanwhile in Malaysia, a total of 16,325 fatalities from ischemic heart disease were reported by the state as in Figure 1 [3]. The data in Figure 1 should spark us as it is crucial to acknowledge the heart problems as soon as possible to begin treatment with therapy and medications.

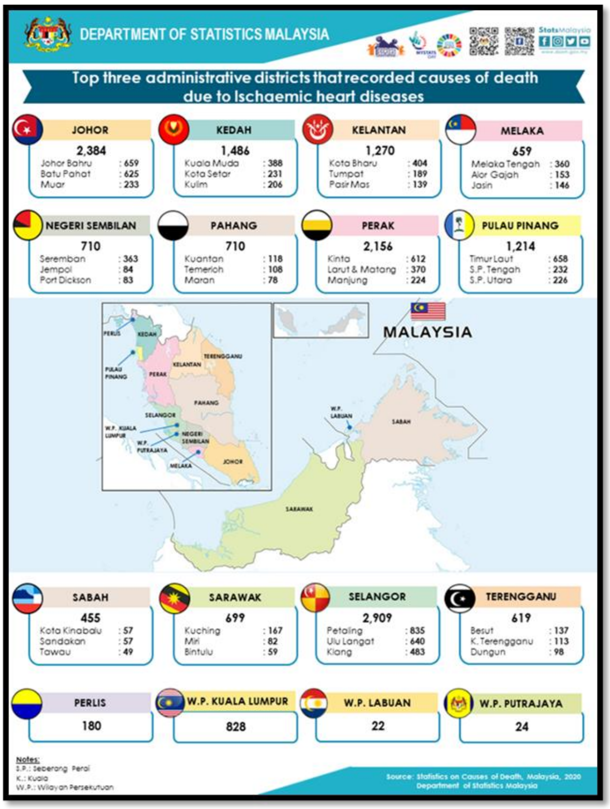

Fig. 1. Record of death due to ischemic heart disease in Malaysia (2020) [3]

Besides, the availability of current technologies such as prediction systems can help in providing early predictions or estimations about the condition of our hearts, on the other hand, can reduce the risk of fatality due to heart failure. An electrocardiogram or known as ECG screening is one example of a technology that is currently available to evaluate the health of a person's heart. This issue has also prompted the interest of many researchers in developing prediction models for identifying heart problems. It involves the use of machine learning and data mining, both of which are growing popular in such systems.

Reflecting on the dangerous scenario, that eventually goes on indefinitely, this project has been introduced that would allow people to check their heart condition early to prevent the terrible situation from happening if no action is taken. HealthyHeart dashboard focused on the prediction of heart conditions whether the user might have heart disease or not. Certain parameters and risk factors have been used as guidance in developing predictive models. The intended parameters are age, sex, chest pain, cholesterol level, and other domain associated [4].

2. Literature Review

2.1 Heart Disease

According to Sazlina *et al.,* [5], heart disease has become the leading cause of mortality worldwide over the past few decades. Based on the data from the WHO [5,6], over 31% of global deaths are caused by CVD, which is primarily coronary artery disease and stroke. In Malaysia, over 13.2% died in 2016 because of CVD, followed by a stroke at 6.9%. Meanwhile, in more developed countries like Ukraine and Russia, each of them has about 718 and 654 deaths per 100,000 population, respectively, while South Korea and Japan have the lowest, with 36.5 and 47.0 deaths per 100,000 population, respectively [7].

Figure 2 shows the proportional rate of mortality due to NCDs. The higher rate of it is cardiovascular disease. About 35% of the total deaths are cardiovascular disease, 17% for communicable, maternal, perinatal, and nutritional conditions, 16% for cancers, 16% for other NCDs, and 9 % for injuries. 4% for chronic respiratory diseases, and another 3% for diabetes. Thus, a system that can help in raising awareness and making an early prediction is required [20]. This is where HealthyHeart comes to the rescue.

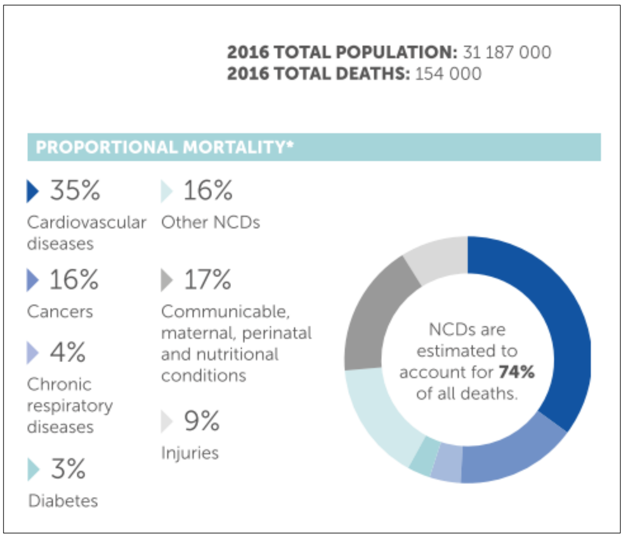

Fig. 2. Proportional mortality of noncommunicable diseases (NCDs) [6]

2.2 Big Data Concept

The volume of data produced by people, machines, and businesses all across the world has grown significantly over the years. According to Wahyuningsih [8], big data can also be defined as the extensive amount of data that is complex and cannot be processed or managed by traditional data processing systems. In recent years, big data has become famous in many sectors. It includes industries such as education, information technology, and government. The significant trend in data growth is due to the record from the Internet of Things and the digitization of all offline information such as medical and education. Big data has become essential for today's industries since most of the sectors and industries have used it to store and analyze their data.

2.3 Big Data and Decision Making

According to Ali *et al.,* [9], big data has transformed many lives. It plays a vital role in deciding various sectors. New opportunities in the domain of visualization also can be created as it can innovate ideas to solve big data problems using visuals. With big data, an interesting pattern can be created based on the data analysis. The visuals from the data representations then can be used to help our brain to digest the meaning of the data patterns and take decisions accordingly. It makes complex things easier to be conveyed with ease and faster. Figure 3 below shows the benefits of data visualization where decision-making has improved by 77% with the use of big data visualization. Thus, this HealthyHeart project has been visualized in dashboard form to capture the user's attention to easily understand the presented data about the health condition.

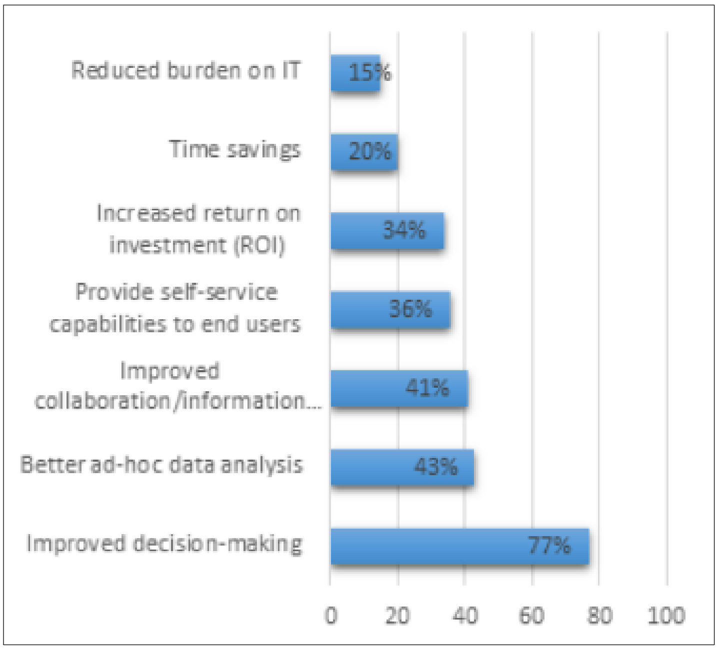

Fig. 3. Benefits of big data visualization [9]

2.4 Big Data Application in Healthcare Area

The presence of big data analytics has a significant impact on the way people experience health care, allowing for more personalized approaches to be achieved more frequently [10]. More exciting ideas are assumed to be created by the innovators due to the rapid growth in data through EHR, medication refill records, insurance reports, genomics, telemedicine, sensor data, and many more [10]. The traditional way does not utilize the full benefits of the insights as well as big data does with many new enhancements [10].

Computerized health information technology has shrunk previously insurmountable distances, enabling the collection of digital, text, voice, and image data that helps decision-making [11]. It is also useful in identifying the training needs, continuous research studies, and life and crucial for data management in health industries such as management systems, telemedicine, and medical and administrative decision support systems [11,12].

Big data is also beneficial for healthcare providers since it can be utilized to assess user status via a parallel computing approach to validate personalized medical data applications [11,13]. To make it happen, integration of data pools is needed to create opportunities. Table 1 shows the main data needed in healthcare.

Table 1 explains one of the data pool types that has been applied to the HealthyHeart project is clinical data. To generate the predictive model, clinical data is required to ensure the system learns the pattern and the behavior of the dataset before a prediction can be made. The clinical dataset that has been used for HealthyHeart is the real dataset from five different places, such as Cleveland for 303 observations, Hungarian for 294 observations, Switzerland for 123 observations, Long Beach, VA for 200 observations, and Stalog (Heart) for 270 observations. This dataset combines over 11 common factors as well as the outcome result.

Table 1
Main data pools needed in healthcare [10]

Primary Data Pools	Description
Clinical data	Owners: Providers Example dataset: medical images, electronic medical records (EMR)
Activity (claims) and cost data	Owners: Providers, Payors Example dataset: utilization of care, cost estimates
Pharmaceutical R&D data	Owners: Academia, Pharmaceutical companies Example dataset: Clinical trials, high-throughput-screening libraries
Patient behavior and sentiment data	Owners: Consumers and stakeholders outside health care (e.g., apparel, retail) Example dataset: retail purchase history, patient behaviors, and preferences, and exercise data captured in running shoes.

Based on the research from Alexander and Wang [10] and Ismail *et al.,* [11], the issues of big data in the health area are crucial for decision-making. Thus, HealthyHeart is suitable for users to decide based on the prediction result of their heart risk condition and heart insight shown in this HealthyHeart dashboard. Users also can view the risk factors and be aware of the heart disease factors for early prevention action.

2.5 Heart Disease Prediction Model System Architecture

The prediction system consists of a few steps before it can proceed with the further process. Data cleaning and filtering need to be applied to the dataset to sort out those irrelevant data [14]. Heart disease prediction system (EHDPS) can be used as a data mining technique where it can be classified into three steps as mentioned by Jothi *et al.,* [4]: Data collection, Data pre-processing, and the classification of data. On another hand, Singh *et al.,* [14] also proposed the same three models as cited in Jothi *et al.,* [4]. The modules include Data collection or Information Gathering, Data pre-processing or Handling Missing Data, and Classification and Prediction.

Figure 4 explains the system architecture of the predictive model for heart disease where firstly clinical dataset would be obtained from medical expertise, then the data would go through a few phases which are data collection, data pre-processing, and classification. By using appropriate data mining techniques and tools, the dataset then can be used as a prediction model which is a heart disease predictive model.

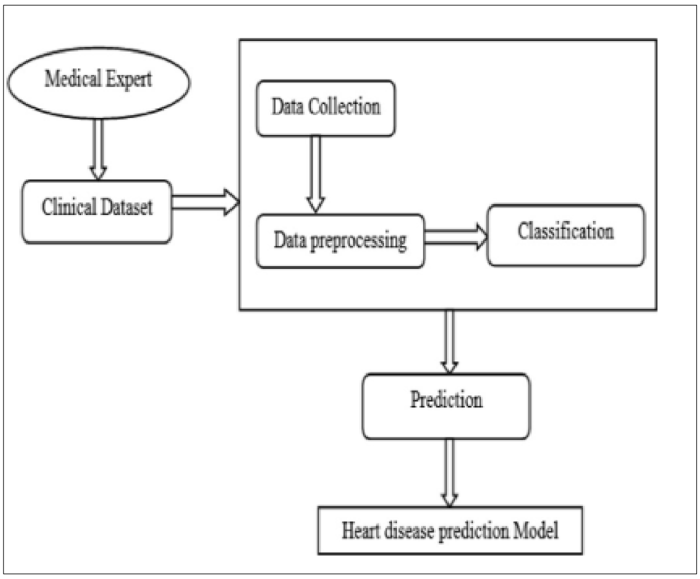

Fig. 4. System architecture of heart disease prediction model [4]

2.6 Type of Algorithm Used

According to Theerthagiri [15], Gradient Boosting (GB) is efficient and contemporary, where it uses gradient boosting learning as a method. Based on the research, gradient boosting shows a great performance as an ensemble classifier for generalization. To be added, GB can have regularization terms to determine the complexity of the model and prevent any overfitting.

Figure 5 shows the workflow of Gradient Boosting (GB) classifier algorithms which started with loading the heart dataset, data pre-processing, feature selection and ensemble-based classification, performance evaluation, error analysis, AUROC analysis, and lastly classifying whether the patients have heart disease or not.

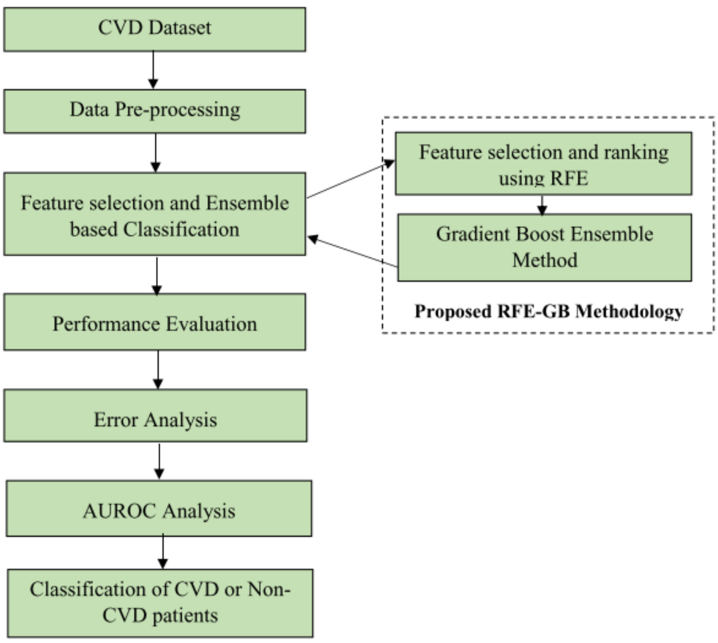

Fig. 5. Workflow of gradient boosting algorithms [15]

2.7 Type of Environment Used

Liang *et al.,* [16] stated that Flask can be defined as the micro web framework that uses Python-based which has strong compatibility and high scalability. Based on Figure 6, the flask framework uses the Werkzeug toolkit to support the integrations of the routes and the jinja2 template engine with sandbox execution functions. High-efficiency websites that require a data analysis platform can be done flexibly with Flask [16].

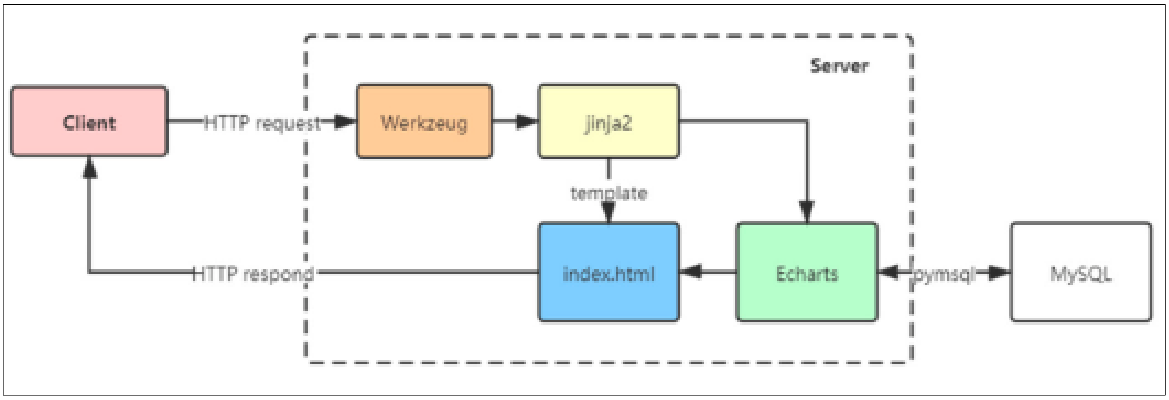

Fig. 6. Webpage response flow chart under the Flask framework [16]

Liang *et al.,* [16] also mentioned that Flask's basic mode is to assign and bind different view functions to different access paths through the route decorator which helps in saving steps of homepage navigation, as well as making sure that a unique URL is generated. With this, users can directly access the interface that they want. For example, to access HealthyHeart predictive model, the user can simply use the URL "/predictForm" on localhost or "https://healthyrate.herokuapp.com/predictForm" for online view.

Based on this research, Flask is being chosen to be the environment for the HealthyHeart project as it allows the combination of an application layer, service layer, and data layer as shown in Figure 7 where HealthyHeart has an application layer that shows heart insight and easy flow for human-computer interactions, meanwhile for service layer in HealthyHeart is data service: HTTP that connect to front-end framework using Bootstrap, HTML, CSS, and Flask as the back-end framework environment. Lastly, for the data layer in HealthyHeart, it inherits the local dataset of the heart disease details that have been uploaded into the Flask which can be connected to data analysis for machine learning [19].

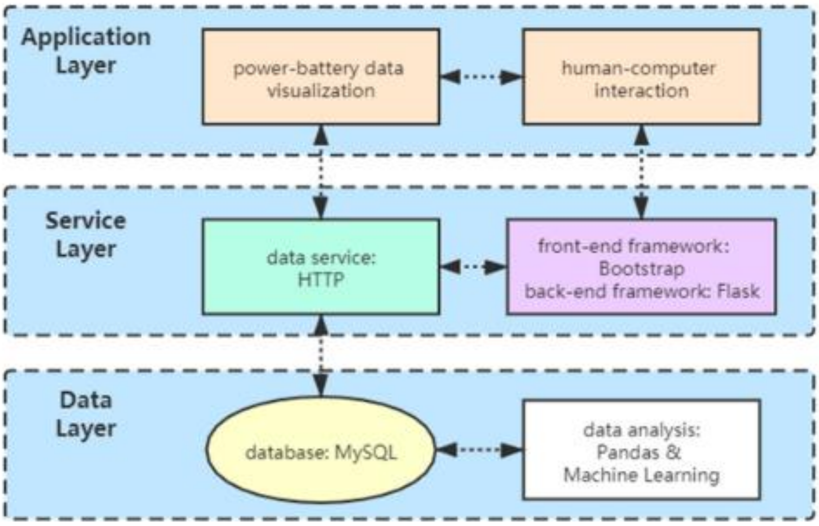

Fig. 7. Architecture diagram using Flask

3. Methods

An Agile diagram and an OSEMN as shown in Figure 8 are the data pipeline model that has been chosen for this research. This phase is very crucial in making sure that all the preparations like data, software, hardware, flow, and tools that are required are prepared accordingly on time.

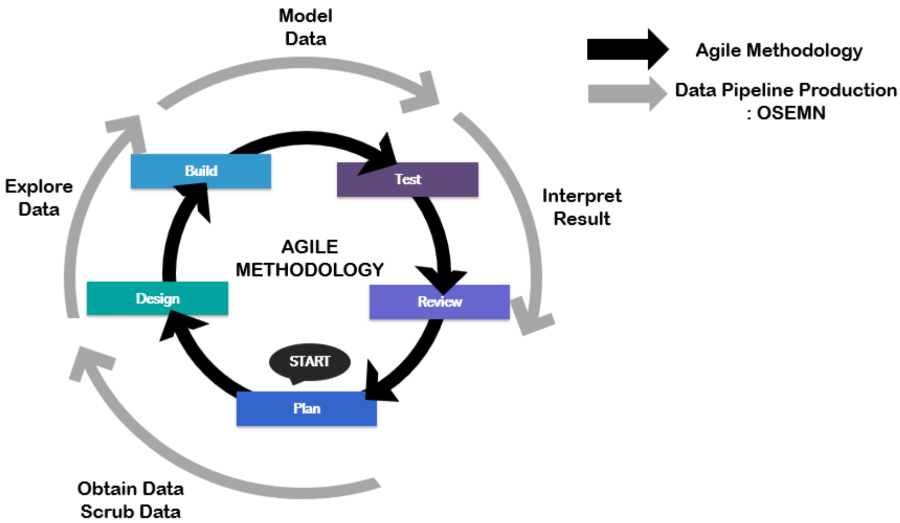

Fig. 8. Process diagram of HealthyHeart using Agile Methodology and OSEMN model [17]

The first step in developing the HealthyHeart dashboard is the planning phase, where all the data requirements are gathered. Understanding the main problem is essential before starting any development because being deprived of understanding the core problem might lead to failure in building a system. As for HealthyHeart, the problem has been defined first before moving into the next steps.

For the HealthyHeart project, the dataset used is from the UCI Machine Learning Repository [14]. The dataset consists of a combination of five heart datasets with 11 parameters where it has been used to recognize the existence of heart disease. The five datasets are from Cleveland (303 observations), Hungarian (294 observations), Switzerland (123 observations), Long Beach VA (200 observations), and Stalog heart datasets (270 observations). The parameters used in this research from Singh *et al.,* [14] as shown in Figure 9 below except for a few parameters such as the number of major vessels, defect type, and obesity. It is due to the limitation of the dataset provided. The data have then been cleaned and classified into the data model for machine learning to build the predictive model. Few algorithms have been chosen and tested for the accuracy level before the predictive model can be produced. After the algorithms are chosen, the predictive model is then saved as the object pickle and integrated into the Flask environment.

S no	Parameters	Parameter description	Values
1	age	Age in years	Continuous
2	sex	Male or female	1= male 0= female
3	thestbps	Resting blood pressure	Continuous value in mmHg
4	cp	Chest pain type	1= typical type 1 2= typical type angina 3= non-angina pain 4= asymptomatic
5	chol	Serum cholesterol	Continuous value in mm/dL
6	fbs	Fasting blood sugar	1≥120 mg/dL 0≤120 mg/dL
7	restecg	Resting electrographic results	0= normal 1= having ST-T wave abnormal 2= left ventricular hypertrophy
8	thalach	Maximum heart rate achieved	Continuous value
9	old peak	ST depression induced by exercise relative to rest	Continuous value
10	exang	Exercise induced angina	0= no 1= yes
11	ca	Number of major vessels colored by fluoroscopy	0–3 value
12	slope	Slope of the peak exercise ST segment	1= unsloping 2= flat 3= downsloping
13	thal	Defect type	3= normal 6= fixed 7= reversible defect
14	obes	Obesity	1= yes 0= no
15	num	Diagnosis of heart disease	0%≤50% 1%>50%

Fig. 9. Description of 15 parameters used by Singh *et al.,* [14]

The second phase of HealthyHeart development is the design phase. During this phase, anything involving the user interface (UI) is selected carefully, because a good user interface helps the interaction between the user and the system run smoothly. The principles of human-computer interaction are crucial during the design process. Thus, the high-fidelity storyboard has been completed to represent the flow of the HealthyHeart dashboard and the predictive model before continuing to the development stage.

The third phase is the building phase. In the building phase, all the work of executing the HealthyHeart dashboard starts here. During this phase, the raw dataset that we got earlier has gone through data processing where it has been cleaned and processed. Data processing and cleaning are very important in data mining as they can reduce the rate of inaccuracy of the data. According to Gupta and Chandra [18], one of the most important tasks is to discover patterns in large datasets using data mining. Its main function of it is to eliminate discrimination and outliers from the existing data. Gupta and Chandra [18] stated that one of the important tasks to discover the pattern from large data is by using data mining. HealthyHeart has gone through data cleaning such as removing null values and removing duplicate rows.

4. Results and Discussion

4.1 Identification of Data and System Requirements

4.1.1 Data collection

The dataset used is in this research retrieved from UCI Machine Learning Repository which is combined and published on Kaggle. The dataset consists of a combination of five heart datasets with 11 parameters where it has been used to recognize the existence of heart disease. The five datasets are from Cleveland (303 observations), Hungarian (294 observations), Switzerland (123 observations), Long Beach VA (200 observations), and Stalog heart datasets (270 observations) and it is in Excel (CSV) format. The parameters used in this research from Singh *et al.,* [14] as shown in Figure 9 except for a few parameters such as the number of major vessels, defect type, and obesity. It is due to the limitation of the dataset provided. However, the dataset used in this research is way bigger than what is in research done by Singh *et al.,* [14] which is 918 data compared to the existing study is only 303 data. Based on this research also, parameters to determine heart disease have been identified.

Each of the datasets contains the information like age of the patient, gender, chest pain type, resting blood pressure, cholesterol level, fasting blood sugar level, resting electrocardiogram result, maximum heart rate achieved, exercise-induced angina, ST old peak value, the slope of the peak exercise ST segment, and the result either the patient has heart disease or not. This data is very useful to determine and predict the existence of heart disease in the patient.

4.1.2 Data pre-processing

As for the second step which is data pre-processing, data that has been obtained needs to be cleaned by identifying the required fields and eliminating all the duplicates as shown in Figure 10. Next, all the missing values would be assigned to the domain value or replaced with the mean mode method (Figure 11) [4,14]. Based on this research, the dataset obtained for the HealthyHeart project has undergone a few steps in data pre-processing. It includes steps such as removing null values and duplicate rows. It is to ensure that the data shown is clean and the prediction is accurate.

```
In [7]: #2.Check duplicate rows using DataFrame.drop_duplicates() and keep first duplicate row
        df = df.drop_duplicates(keep='first')

In [8]: #check rows AFTER drop duplicate rows (No duplicate rows)
        df.shape

Out[8]: (918, 12)
```

Fig. 10. Check duplicate rows

```
In [6]: #1. Check null values
        df.isnull().sum()

Out[6]: Age               0
        Sex               0
        ChestPainType     0
        RestingBP         0
        Cholesterol       0
        FastingBS         0
        RestingECG        0
        MaxHR             0
        ExerciseAngina    0
        Oldpeak           0
        ST_Slope          0
        target            0
        dtype: int64
```

Fig. 11. Check null values

4.1.3 Classification and prediction

For the predictive model, 11 parameters are used to define the status and the pattern of the disease. The parameters like age, gender, chest pain type (CP), cholesterol, and some other associated values as tabulated in Figure 12. Other parameter associates like blocked pressure level and electrocardiographic (ECG) results can also be taken into consideration [4]. However, the prediction still can be done as the dataset provides a target or result outcome so that machine learning can be made based on the factors and parameters provided in the dataset.

```
In [3]: #checking the row and col
        df.shape

Out[3]: (918, 12)

In [4]: #getting the info of the dataframe
        df.info()

        <class 'pandas.core.frame.DataFrame'>
        RangeIndex: 918 entries, 0 to 917
        Data columns (total 12 columns):
         #   Column          Non-Null Count  Dtype
        ---  ------          --------------  -----
         0   Age             918 non-null    int64
         1   Sex             918 non-null    int64
         2   ChestPainType   918 non-null    int64
         3   RestingBP       918 non-null    int64
         4   Cholesterol     918 non-null    int64
         5   FastingBS       918 non-null    int64
         6   RestingECG      918 non-null    int64
         7   MaxHR           918 non-null    int64
         8   ExerciseAngina  918 non-null    int64
         9   Oldpeak         918 non-null    float64
         10  ST_Slope        918 non-null    int64
         11  target          918 non-null    int64
        dtypes: float64(1), int64(11)
        memory usage: 86.2 KB
```

Fig. 12. 11 parameters used

After transforming raw data into meaningful information, then the process of data extraction and selection has been applied to select a dataset of related fields or guessers to build the components. Other names for data extraction and selection are variable selection or attribute selection [4]. Then, it is time to prepare for the data model of machine learning. It starts with initializing the target and the features within the dataset. Next, the dataset is split into 2 sets which are the training set and the testing set as shown in Figure 13.

Right after that, the data is trained and evaluated as in Figure 13. A few algorithms such as RandomForestClassifier, AdaBoostClassifier, and GradientBoostingClassifier are used to determine which algorithms perform best for the prediction as in Figure 14. The results are then shown in Figure 15 the classification report and confusion matrix.

```
In [54]: # Train and evaluate model
def fit_eval_model(model, train_features, y_train, test_features, y_test):

    """
    Function: train and evaluate a machine learning classifier.
    Args:
      model: machine learning classifier
      train_features: train data extracted features
      y_train: train data lables
      test_features: train data extracted features
      y_test: train data lables
    Return:
      results(dictionary): a dictionary of classification report
    """
    results = {}

    # Train the model
    model.fit(train_features, y_train)

    # Test the model
    train_predicted = model.predict(train_features)
    test_predicted = model.predict(test_features)

     # Classification report and Confusion Matrix
    results['classification_report'] = classification_report(y_test, test_predicted)
    results['confusion_matrix'] = confusion_matrix(y_test, test_predicted)

    return results
```

Fig. 13. Train and evaluate model

```
In [55]: # Initialize the models
sv = SVC(random_state = 1)
rf = RandomForestClassifier(random_state = 1)
ab = AdaBoostClassifier(random_state = 1)
gb = GradientBoostingClassifier(random_state = 1)


# Fit and evaluate models
results = {}
for cls in [sv, rf, ab, gb]:
    cls_name = cls.__class__.__name__
    results[cls_name] = {}
    results[cls_name] = fit_eval_model(cls, X_train, y_train, X_test, y_test)
```

```
In [56]: # Print classifiers results
for result in results:
    print (result)
    print()
    for i in results[result]:
        print (i, ':')
        print(results[result][i])
        print()
    print ('-----')
    print()
```

Fig. 14. Initialization and classify the algorithms

```
RandomForestClassifier

classification_report :
              precision    recall  f1-score   support

           0       0.83      0.78      0.81        77
           1       0.85      0.89      0.87       107

    accuracy                           0.84       184
   macro avg       0.84      0.83      0.84       184
weighted avg       0.84      0.84      0.84       184


confusion_matrix :
[[60 17]
 [12 95]]

-----
```

```
AdaBoostClassifier

classification_report :
              precision    recall  f1-score   support

           0       0.83      0.78      0.81        77
           1       0.85      0.89      0.87       107

    accuracy                           0.84       184
   macro avg       0.84      0.83      0.84       184
weighted avg       0.84      0.84      0.84       184


confusion_matrix :
[[60 17]
 [12 95]]
```

```
GradientBoostingClassifier

classification_report :
              precision    recall  f1-score   support

           0       0.84      0.81      0.82        77
           1       0.86      0.89      0.88       107

    accuracy                           0.85       184
   macro avg       0.85      0.85      0.85       184
weighted avg       0.85      0.85      0.85       184


confusion_matrix :
[[62 15]
 [12 95]]

-----
```

Fig. 15. Classifiers result for RandomForestClassifier, AdaBoostClassifier, and GradientBoostingClassifier

Based on the classification report, the Random Forest (RF) classifier has an 84% accuracy level while the GradientBoostingClassifier (GBC) gets the highest accuracy level compared to the other 2 algorithms, hence GBC is chosen as the algorithm for the predictive model. The model is then saved as an object pickle where it will be used later during the system integration process in the Flask environment.

4.2 HealthyHeart System Data Design

This section describes the procedures involved in developing the required datasets for the HealthyHeart dashboard, including how the previously cleaned dataset is used as shown in Figure 16. This project requires a real patient dataset of heart disease. The dataset has been retrieved from UCI Machine Learning Repository which is combined and published on Kaggle. The dataset includes the observation from Cleveland, Hungarian, Switzerland, Long Beach VA, and the Stalog (Heart) dataset in the format of Excel (.csv).

Figure 16 shows the cleaned data from the previous scrubbing processes as mentioned in section 4.1. Parts of the data model that involved the HealthyHeart project are the predictive model using GBC classifier model and data visualization visualized using Tableau software which involved data preparation, data cleaning, data exploration, and data modelling. In addition, in this stage, the design of the user interface has been done before the development stage. A use case diagram (Figure 17) of the HealthyHeart dashboard has also been presented to capture the understanding of the HealthyHeart system flow.

1	Age	Sex	ChestPain	RestingBP	Cholester	FastingBS	RestingEC	MaxHR	ExerciseA	Oldpeak	ST_Slope	target
2	40	1	1	140	289	0	0	172	0	0	0	0
3	49	0	2	160	180	0	0	156	0	1	1	1
4	37	1	1	130	283	0	1	98	0	0	0	0
5	48	0	3	138	214	0	0	108	1	1.5	1	1
6	54	1	2	150	195	0	0	122	0	0	0	0
7	39	1	2	120	339	0	0	170	0	0	0	0
8	45	0	1	130	237	0	0	170	0	0	0	0
9	54	1	1	110	208	0	0	142	0	0	0	0
10	37	1	3	140	207	0	0	130	1	1.5	1	1
11	48	0	1	120	284	0	0	120	0	0	0	0
12	37	0	2	130	211	0	0	142	0	0	0	0
13	58	1	1	136	164	0	1	99	1	2	1	1
14	39	1	1	120	204	0	0	145	0	0	0	0
15	49	1	3	140	234	0	0	140	1	1	1	1
16	42	0	2	115	211	0	1	137	0	0	0	0
17	54	0	1	120	273	0	0	150	0	1.5	1	0
18	38	1	3	110	196	0	0	166	0	0	1	1
19	43	0	1	120	201	0	0	165	0	0	0	0
20	60	1	3	100	248	0	0	125	0	1	1	1
21	36	1	1	120	267	0	0	160	0	3	1	1
22	43	0	0	100	223	0	0	142	0	0	0	0
23	44	1	1	120	184	0	0	142	0	1	1	0

Fig. 16. Cleaned data

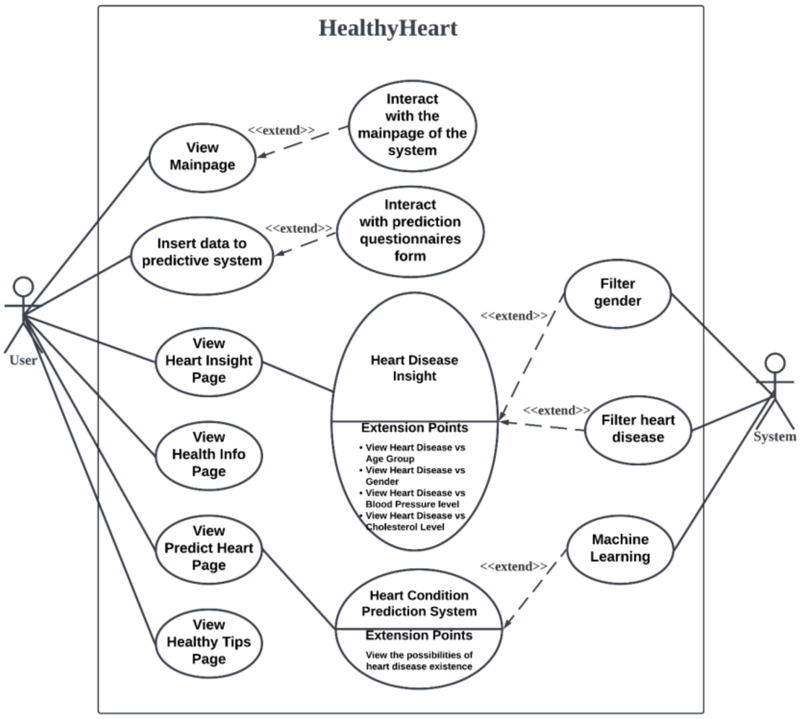

Fig. 17. Use case diagram of HealthyHeart

4.3 Development of HealthyHeart Dashboard with Predictive Model

The third objective is the development of the HealthyHeart dashboard. The design of the interface is made by Tableau, whereas the predictive model is made by using HTML and CSS. Figure 18 shows the dashboard of HealthyHeart.

Figure 18 illustrates the HealthyHeart dashboard which consists of multiple pages such as Heart Insight, Health Info, Predict Heart, and Healthy Tips. Users can view general information about the heart through Heart Insight and Health Info page. Furthermore, users can predict their heart condition when they go to predict heart page by including a few related inputs. From the inputs, users can know whether they have a healthy heart or not. Finally, healthy heart tips are provided to users to aware them of taking care of their heart so that they will not be the ones who contribute to heart disease cases.

For Figure 19, the prediction system is embedded into the dashboard system with a hyperlink button. There are a few quick tips for the users to read before they can jump into the prediction system.

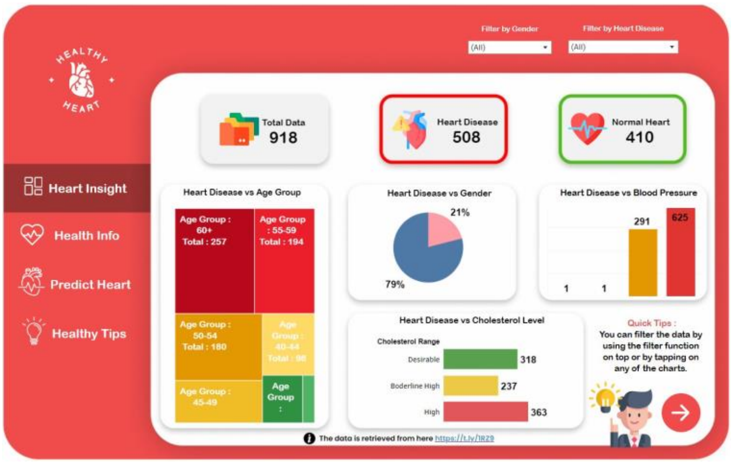

Fig. 18. HealthyHeart dashboard

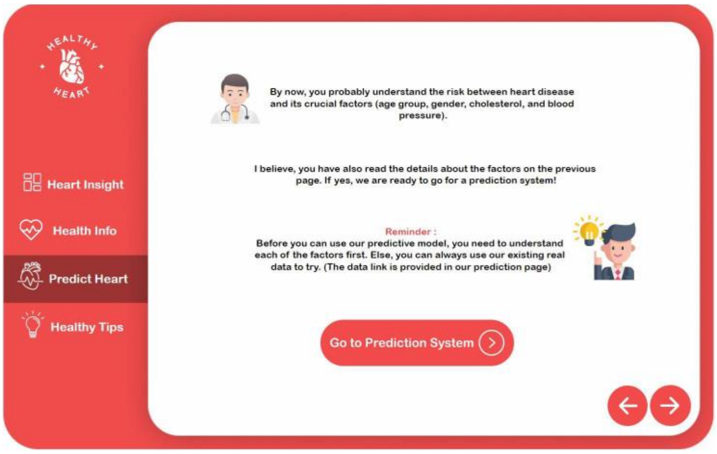

Fig. 19. Prediction heart page

Figure 20 indicates the predictive system using the GBC classifier model, where in here the users can put all the values of the attributes needed before the system can do the prediction. They can also try out the existing real data provided on top of the form. The result page shows when the user's heart is normal. However, there are disclaimers stating that the prediction is based on machine learning only and it does not replace any medical officers' advice.

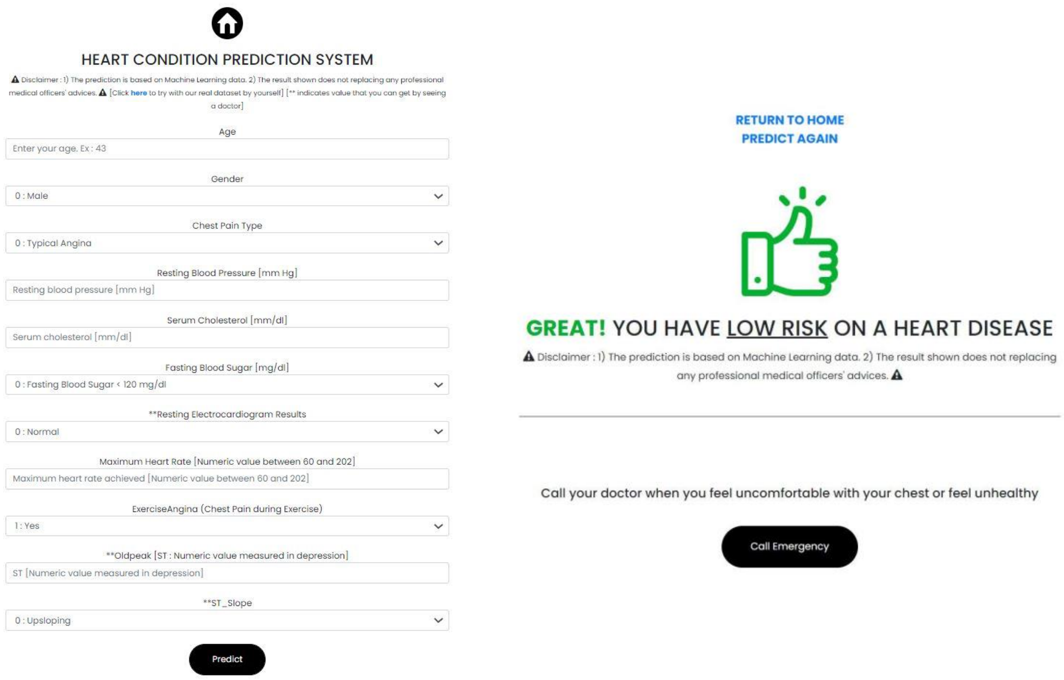

Fig. 20. Questionnaire for prediction system and the results

5. Conclusions

In a nutshell, the objective of the system has been met as all the functionality is working well. The HealthyHeart dashboard was carefully developed according to all of the data requirements gathered throughout the whole year. Many people can be benefited and be educated from this dashboard. Civilians can learn more and increase their awareness regarding heart disease by using the predictive model and also data visualization within the HealthyHeart dashboard. They can identify the risk factors that contribute to heart disease well better and learn more about it. Besides, medical experts also can use the predictive model on their patients. Therefore, precautions can be taken seriously, and early detection also can be made which may help to lower the risk of mortality with heart disease. Besides, by having more data, machine learning can learn better and predict the results with higher accuracy levels.

Acknowledgment
The authors would like to thank the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Malaysia, and iKohza software engineering, University Technology Malaysia for their endless support in completing this research.

References
[1] Benjamin, Emelia J., Paul Muntner, Alvaro Alonso, Marcio S. Bittencourt, Clifton W. Callaway, April P. Carson, Alanna M. Chamberlain et al. "Heart disease and stroke statistics—2019 update: a report from the American Heart Association." *Circulation* 139, no. 10 (2019): e56-e528.
[2] Virani, Salim S., Alvaro Alonso, Emelia J. Benjamin, Marcio S. Bittencourt, Clifton W. Callaway, April P. Carson, Alanna M. Chamberlain et al. "Heart disease and stroke statistics—2020 update: a report from the American Heart Association." *Circulation* 141, no. 9 (2020): e139-e596.
[3] DOSM. "Department of Statistics Malaysia. " (2000)
[4] Jothi, K. Arul, S. Subburam, V. Umadevi, and K. Hemavathy. "WITHDRAWN: Heart disease prediction system using machine learning." (2021).

[5] Sazlina, Shariff-Ghazali, Ping Yein Lee, Yoke Mun Chan, Mohamad Shariff A. Hamid, and Ngiap Chuan Tan. "The prevalence and factors associated with sarcopenia among community living elderly with type 2 diabetes mellitus in primary care clinics in Malaysia." *PLoS One* 15, no. 5 (2020): e0233299. https://doi.org/10.1371/journal.pone.0233299

[6] World Health Organization. "World Health Organization-WHO." (2000).

[7] Seong, Ang Choon, and Chan Kok Meng John. "A review of coronary artery disease research in Malaysia." *Med J Malaysia* 71, no. Supplement 1 (2016): 46.

[8] Wahyuningsih, Tri. "Problems, Challenges, and Opportunities Visualization on Big Data." *Journal of Applied Data Sciences* 1, no. 1 (2020): 20-28. https://doi.org/10.47738/jads.v1i1.8

[9] Ali, Syed Mohd, Noopur Gupta, Gopal Krishna Nayak, and Rakesh Kumar Lenka. "Big data visualization: Tools and challenges." In *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, pp. 656-660. IEEE, 2016. https://doi.org/10.1109/IC3I.2016.7918044

[10] Alexander, Cheryl Ann, and Lidong Wang. "Big data analytics in heart attack prediction." *J Nurs Care* 6, no. 393 (2017): 2167-1168. https://doi.org/10.4172/2167-1168.1000393

[11] Ismail, Ahmed, Samir Abdlerazek, and I. M. El-Henawy. "Big data analytics in heart diseases prediction." *Journal of Theoretical and Applied Information Technology* 98, no. 11 (2020): 15-19.

[12] Kamran, M., and A. Javed. "A survey of recommender systems and their application in healthcare." *University of Engineering and Technology Taxila. Technical Journal* 20, no. 4 (2015): 111.

[13] Dumka, Ankur, and Anushree Sah. "Smart ambulance system using concept of big data and internet of things." In *Healthcare data analytics and management*, pp. 155-176. Academic Press, 2019. https://doi.org/10.1016/B978-0-12-815368-0.00006-3

[14] Singh, Poornima, Sanjay Singh, and Gayatri S. Pandi-Jain. "Effective heart disease prediction system using data mining techniques." *International journal of nanomedicine* 13, no. sup1 (2018): 121-124. https://doi.org/10.2147/IJN.S124998

[15] Theerthagiri, Prasannavenkatesan. "Predictive analysis of cardiovascular disease using gradient boosting based learning and recursive feature elimination technique." *Intelligent Systems with Applications* 16 (2022): 200121. https://doi.org/10.1016/j.iswa.2022.200121

[16] Liang, Zuyi, Zongwei Liang, Yubin Zheng, Beichen Liang, and Linfeng Zheng. "Data Analysis and visualization platform design for batteries using Flask-based Python Web Service." *World Electric Vehicle Journal* 12, no. 4 (2021): 187. https://doi.org/10.3390/wevj12040187

[17] Nantasenamat, Chanin. "The Data Science Process." Experfy Insights, August 26, 2020. https://resources.experfy.com/bigdata-cloud/the-data-science-process/.

[18] Gupta, Manoj Kumar, and Pravin Chandra. "A comprehensive survey of data mining." *International Journal of Information Technology* 12, no. 4 (2020): 1243-1257. https://doi.org/10.1007/s41870-020-00427-7

[19] Zamani, Hadi, and Muhamad Kamal Mohammed Amin. "Classification of phishing websites using machine learning techniques." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 5, no. 2 (2016): 12-19.

[20] Ojo, Emmanuel, Olayimika Mercy Oguns, and James Eromi Odeiani. "A Study on Prevalence of Hypertension and Challenges of High Cost of Medications in General Out Patient Department (GOPD) FMC, a Tertiary Care Medical Centre, at Lokoja, North Central Nigeria." *Journal of Advanced Research in Applied Sciences and Engineering Technology* 12, no. 1 (2018): 32-40.