DAYANANDA SAGAR ACADEMY OF TECHNOLOGY & MANAGEMENT
Opp. Art of Living, Udayapura, Kanakapura Road, Bangalore- 560082
Affiliated to Visvesvaraya Technological University, Belagavi and Approved by AICTE, New Delhi
DEPARTMENT OF INFORMATION SCIENCE & ENGINEERING
Accredited by NBA, New Delhi

Project Phase 1(18CSP77) Report

on

Heart disease prediction using Machine Learning

Submitted by

1DT19IS095	**PRAKRUTHI HR**
1DT19IS079	**MOULYA G**
1DT19IS086	**NAYANA SAGAR**
1DT19IS101	**PRIYANSHU SINGH**

Under the Guidance of
Dr.ThiruKrishna JT,
Associate Professor,
Dept of ISE,
DSATM

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

According to the World Health Organization (WHO), heart attacks account for 80 percent of the overall fatalities nationwide. Each person has a varied blood pressure and heart rate, which range between 60 to 100 beats per minute for pulse rates and 120/80 to 140/90 for blood pressure.

It is difficult to identify heart disease because of several contributory risk factors such as diabetes, high blood pressure, high cholesterol, abnormal pulse rate and many other factors. Various techniques in data mining and neural networks have been employed to find out the severity of heart disease among humans. The severity of the disease is classified based on various methods like K -Nearest Neighbor Algorithm (KNN), Decision Trees (DT), Genetic algorithm (GA), and Naive Bayes (NB). The nature of heart disease is complex and hence, the disease must be handled carefully. Not doing so may affect the heart or cause premature death. The perspective of medical science and data mining are used for discovering various sorts of metabolic syndromes. Data mining with classification plays a significant role in the prediction of heart disease and data investigation.

In recent times, Heart Disease prediction is one of the most complicated tasks in medical field. In the modern era, approximately one person dies per minute due to heart disease. Data science plays a crucial role in processing huge amount of data in the field of healthcare. As heart disease prediction is a complex task, there is a need to automate the prediction process to avoid risks associated with it and alert the patient well in advance. This paper makes use of heart disease dataset available in UCI machine learning repository. The proposed work predicts the chances of Heart Disease and classifies patient's risk level by implementing different data mining techniques such as Naive Bayes, Decision Tree, Logistic Regression and Random Forest. Thus, this paper presents a comparative study by analysing the performance of different machine learning algorithms. The trial results verify that Random Forest algorithm has achieved the highest accuracy of 90.16% compared to other ML algorithms implemented.

Arrhythmia	The heart beat is improper whether it may irregular, too slow or too fast.
Cardiac arrest	An unexpected loss of heart function, consciousness and breathing occur suddenly.
Congestive heart failure	The heart does not pump blood as well as it should, it is the condition of chronic.
Congenital heart disease	The heart's abnormality which develops before birth.
Coronary artery disease	The heart's major blood vessels can damage or any disease occurs in the blood vessels.
High Blood Pressure	It has a condition that the force of the blood against the artery walls is too high.
Peripheral artery disease	The narrowed blood vessels which reduce flow of blood in the limbs, is the circulatory condition.
Stroke	Interruption of blood supply occur damage to the brain.

Table 1.1 showing different types of heart diseases

CHAPTER 2
PROBLEM STATEMENT

Literature Survey Research Statement
Certain Investigation on heart disease prediction using Machine Learning

Preview

In this survey we have found different research that have been done by different researchers.The different researchers will tell the different methodologies and their accuracy according to their findings.We have referred 24 papers.Research objectives **describe what our research is trying to achieve and explain why we are pursuing it**. They summarize the approach and purpose of your project and help to focus our research. Our objectives should appear in the introduction of your research paper, at the end of our problem statement.

Objectives of the research

- Different algorithms for prediction of heart disease
- Motivate our work by explaining its social value.
- Provide background context for our readers.
- Compare our findings with other research.
- Justify our experimental method.
- Communicate our project's scientific novelty by framing a research gap.

CHAPTER 3
PROBLEM OBJECTIVES AND SCOPE

The problem objective **specifies the overall goal of the problem**.Before we begin troubleshooting a problem, we must first define the scope of our problem. When defining the scope , we need to **identify what is working and what is not working**. Sometimes it is useful to identify another machine that is working as you expect.

Objectives

- Improve cardiovascular health and quality of life through prevention, detection, and treatment of risk factors for heart attack and stroke
- Early identification and treatment of heart attacks and strokes
- prevention of repeat cardiovascular events
- reduction in deaths from cardiovascular disease

Scope

- Can be used on large scale by hospitals in future.
- More self checks can be done in early stage of symptoms.
- Can utilize different subsets and find the best combinations to increase the dataset's predictive accuracy

CHAPTER 4
LITERATURE SURVEY AND THEORY

Literature Survey have been divided into following parts:

1. Abstract & Introduction
2. Related work
3. Machine learning Algorithms and Model
4. Conclusion and References

4.1 Abstract & Introduction

One of the leading causes of mortality is cardiovascular disease. Since the health system produces a significant amount of data, detecting cardiovascular problems is becoming even more crucial. Internet - of - things healthcare systems provide a tough challenge. Machine learning is essential for making correct disease predictions. There has been extensive work in this domain, yet they have not effectively grasped the real potential of machine learning strategies in predicting risk in patients because they have not utilized large quantities of information. In this study, we suggest a unique method for enhancing the accuracy of coronary heart disease diagnosis by identifying significant features using machine learning strategies. Various feature groupings and many well-known classifications techniques have been employed to develop the prediction system. The main goals of the planned study are to improve feature selection and minimize the number of traits while producing improved outcomes. In this work, a better search optimization algorithm with a conceptual methodology is applied to recognize defining factors of cardiovascular diseases. The proposed technique can also be immediately put into practice in the medical world to detect heart disease.

According to the World Health Organization (WHO), heart attacks account for 80 percent of the overall fatalities nationwide. Each person has a varied blood pressure and heart rate, which range between 60 to 100 beats per minute for pulse rates and 120/80 to 140/90 for blood pressure. Thrombosis, cardiomyopathy, congestive cardiac failure, arrhythmia, pulmonary disease, sudden cardiac death, valve disease, and congenital heart defects are the numerous forms of cvd.It is usually diagnosed by a doctor after reviewing the individual's medical history, the results of their clinical examination, and any alarming problems. However, it is not possible to identify an individual with HD using the outcomes of this clinical diagnosis.Non-laboratory statistics indicate that a variety of risk factors, such as age, gender, smoking, hypertension, high systolic blood pressure, and high blood pressure treatment, raise the risk and body-mass index, might provide useful information for assessing CVD risk [3]. The majority of these parameters might be viewed as onset indicators and cautions to the person, which can add to the risk score obtained by standard biochemical measures (such as cholesterol values). The adoption of self-assessment questionnaires as a complement to most clinical procedures is due to this.When new modes and medical experts are inaccessible, the diagnosis and treatment of heart disease is really quite challenging.As a consequence, timely identification of heart problems can minimize the number of mortality and enable healthcare professionals to prescribe the most appropriate treatment option. However, a number of unidentified elements even cause expertise in heart disease to wrongly identify the condition. However, it is critical to search for accurate procedures to accurately account for all the uncertain risk factors and detect cardiovascular disease. Scientists have explored a diversity of algorithms using machine learning to determine the best combinations of cardiovascular diseases parameters to aid health care professionals in

improving computer - aided diagnosis methods and the quality of treatments. As it takes expertise and in-depth understanding to anticipate cardiac disease, it is a challenging task. The complexity of the problem is categorized using a variety of techniques, including the K-Nearest Neighbor Algorithm (KNN), Naive Bayes (NB), Decision Trees (DT), and Genetic Algorithm (GA).Recently, the importance of optimization to our everyday life has increased. Population- and evolutionary-based optimization approaches are well-liked and frequently employed in various engineering fields. This growth - based finds the best options out of the numerous available options and provides a setting that is good for problem-solving. A mathematical representation of the system is necessary for the majority of optimization strategies . Making a statistical method for complicated processes might be difficult. The high cost forbids employing the solution time even if the model is established. Due to physical events, it is difficult to create an optimization technique to obtain enough global and local search operators.

Due to the CVDs intricate nature, it must be handled with caution. Failure to follow instructions could increase the risk of death or injury to the heart. Many various types of physiologic illnesses are being revealed according to medical research and machine learning (ML). ML with categorization plays an important role in the detection of HD and data processing .An effective machine learning approach is one that performs well on both seen and hidden samples. This happens because a machine learning approach could just understand the data for training otherwise. Several classifiers were placed through data processing, and it was found that they properly classified 50 percent of the total of the instances on average [16]. Additionally, when a model has been trained and evaluated on a dataset, relevant cross validation techniques and performance evaluation metrics are important.

For testing and training, the machine learning prediction models require adequate data. If balanced datasets are used for model training and testing, machine learning model performance can be improved. Furthermore, by incorporating appropriate and significant elements from the data, the model's prediction skills may be improved. In order to increase performance of the model, data balance and extraction of features are therefore critical.Here, we undertake tests to determine the characteristics of a hybrid machine learning algorithm. The outcomes of the experiment indicate that, in comparison to other methods, hybrid methods have a greater capacity to predict heart disease.

The effectiveness of every classifier in the challenge of classifying cardiovascular disease is examined in this study using four large-scale datasets, including the Cleveland heart disease dataset, Cardiovascular dataset, Framingham cardiovascular disease dataset and Cardio train1 dataset.According to experimental results, this gentle group always does well when compared to certain other classifiers, as indicated by a higher measure, particularly with large datasets like the Cardiovascular and the Cardio-Train1 datasets.

4.2 Related Work

The accuracy of the algorithm for decision trees is the lowest when applying the 10-cross validation technique, coming in at 77.55% when all 13 of the dataset's attributes are utilized. KNN comes in second with an accuracy of 83.16percent of total when k = 9.Nevertheless, the accuracy of the decision tree and SVM with boosting is greater, at 82.17% and 84.81%, respectively.The decision tree technique fared poorly with an accuracy of 42.89% compared to the SVM classifier's accuracy of 85.7655%. SVM achieves an f-measure value of 93.5617% .

In a different research, Gudadhe et al. created a diagnosis system for HD diagnosis utilizing

multi-layer Recurrent neural network and support vector machine (SVM) techniques and achieved accuracy of 80.41%. By combining a neural network with fuzzy logic, Humar et al. [developed the HD recognition system.A technique for diagnosing heart disease based on ML was created by Akil et al. The ANN-DBP algorithm and FS algorithm both performed well. A system for professional medical diagnosis for HD identification was proposed by Palaniappan et al. Artificial Neural Networks (ANN), Decision Trees (DT), and Navies Bays (NB) were used as predictive machine learning models during the development of the system. NB attained 86.12% efficiency, ANN 88.12% accuracy, and DT classifier 80.4% accuracy

In Another research by MOHAN et al. developed a hybrid machine learning strategy for HD detection. He also put forth a novel methodology for choosing important characteristics from the information for machine learning classifiers to use in training and testing. They have an 88.07% classification accuracy rate.

A balancing strategy was established in a small number of research to support decision systems that tackled the mentioned issue.To identify and eliminate outliers and equalize distribution of the data, Fitriyani et al. devised an HD prediction method that uses density-based spatial clustering of applications with noise (DBSCAN) and hybrid synthetic minority over-sampling technique-edited nearest neighbor (SMOTE-ENN). The XGBoost classifier also predicts the patient's status using which an accuracy of 95.9% was achieved using the proposed model [16].To predict cardiac attacks, Waqar et al. suggested using deep learning based on SMOTE. Without feature selection, the author balanced the dataset using the SMOTE technique. A deep neural network was trained and tested to predict the absence and presence of a cardiac arrest using the balanced dataset, and it obtained 96% efficiency .

Technique	Limitations	Advantages	Acc(%)
HD diagnosis using ML classifiers	The Proposed method accuracy is very low.	Computationally less complex.	77
MLP+SVM	Computationally complex.	The performance of the proposed method is high in terms of prediction accuracy.	80.41
ANN+Fuzzy Logic	More execution time required to generate results.	Accuracy is high.	87.4
ANN ensemble based diagnosis system	Computationally complex.	High accuracy.	89.01
HD diagnosis system based on NB, DT and ANN	The NB and DT performance are low.	ANN achieved high performance in term of accuracy	88.12
Three phase technique based on ANN	High computation time.	High accuracy.	88.89
ANN-FUZZY-AHP	Computationally complex.	Achieved high accuracy.	91.1
Relief-Rough set based method for HD detection	Computation time is high.	High accuracy due to selection of appropriate feature for training and testing of the model.	92.32
Hybrid ML method	Low accuracy.	Low computation time.	88.07

Fig 4.1 showing various technique with their accuracy score

METHODOLOGICAL FRAMEWORK

A. *Collection of DataSet*

The dataset is referred as group of connected data which contains data for each instance.An attribute in the dataset contributes to the factor that determines the outcome.Each attributes contributes a certain level to the final outcome but it is not sure that all the attributes have same level of control over the outcome. The dataSet has been obtained from international universities such as University of California Irvine (UCI) 2016–2022,which is recorded from real time observation of the patients suffering from cardioVascular Disease. The original dataset contains 13 attributes,270 subjects and an output class.Each and every property present in UCI dataset play an important role in heart disease prediction.Based on the various physical examination and laboratory tests these dataSet has been accumulated. Based on the survey conducted by the UCI is used to find risky factors of the disease. If a person is classified under the category to be tested further An individual is classified as 'needs further tests then there are many factors due to which the person has to take further tests some of them can be lack of physical fitness,obesity,diabetes,high blood pressure. A person who is classified as non-healthy are the ones who have already had heart-attack or have prolonged chest pain.According to reports that are made on a daily basis the chances of having heart related problems are diabetes, high blood pressure, and they can be facing some symptoms such as chest pain or chest burn and shortness in breathThe other people who did not experience any of the symptoms are classified under the healthy category.

B. *preliminary processing of dataSet*

Pre-processing can be defined as a process that is used to convert raw data into useful format.The major step of preprocessing is formatting the data. Normally, the data we obtain in the raw format contains a lot of missing values,wrong representations ,so this data cannot be used for the machine learning models directly and it has to be cleaned up and made suitable for machine learning models. DataSet plays a major role in creating machine learning models.The main steps of preprocessing models are data cleaning,data integration,data transformation,data reduction.

1. Data Cleaning

This technique is used to remove the missing values ,Noisy data and the inconsistency in the data points.The result of this is to get an accurate output for machine learning models.The problem of missing value occurs when one or more dataset are used to form larger dataset,the most easier way to resolve the issue is to delete those fields before merging.

There is one more technique to fill out the missed values with most probable values and this can be done using logistic regression.

2. Data Integration

The data will be collected from different sources and it has to be integrated for the proper usage and during this integration ,it may lead to several inconsistencies and redundant data.There are three techniques for integrating the data they are data consolidation,data virtualization,data propagation .In data consolidation all the data physically bought together at one place this helps to increase organization and productivity of data integration.data virtualization explains about the viewpoint of the data.In data

propagation with the help of some applications we can transfer the data from one location to another location.

3. Data Reduction

This technique is used to reduce the quantity of data so it helps to reduce the cost associated with it.When we are working with big data this data preprocessing step plays a major role.This technique helps to create faster and more efficient models.

4. Data Transformation

This technique is used to convert the data from one pattern to another pattern.Some of the strategies used for data transformation are Smoothing,Aggregation,normalization,

generalization.In generalization we will convert the data features from low level to high level,Normalization process will convert all variables within some specified unit.Smoothing is used to remove the noise from the data using some algorithms.

The table gives a brief description about chest pain occuring in various age groups along with sex.Here we find the men in the age group of 50-60 are the ones who suffer more from chest pain.Chest pain can again be in four various forms. Chest pain is the most common symptom .

Disease	Female	Male
Stroke	10%	8%
Hypertensive heart disease	2%	1.1%
Rheumatic heart disease	0.15%	0.13%
Cardiovascular and circulatory disease	8%	15%
Endocarditis	0.14%	0.11%
Cardiovascular disease	26.7%	29.2%

Table 4.1. Major Categories under Study.

The dataset from the uci is a multivariate dataset that contains 76 attributes which are related to various blood and body conditions out of which a subset containing 10 attributes are been picked like age,sex,blood pressure,blood glucose level ,cholestrol level, electrocardiogram , heart rate angina induced due to exercise,depression caused to due exercising. The four variations in chest pain are marked as a,b,c,d. They are typical angina,atypical angina, non-anginal pain, asymptomatic respectively.

4.3 Machine learning Algorithms and Model

1. **PRISMA algo**

Without prospectively registering, researchers followed a process that was agreed upon by all authors and adhered to the Preferred Reporting. Items for. Systematic Reviews and. Meta-Analysis (PRISMA) declaration.

The objective of this algorithm is to summarize and assess the best reliable machine-learning method for ischemic heart disease prediction.

PRISMA criteria were followed in conducting this systematic review. Multiple databases, including Science Direct, PubMed, MEDLINE, CINAHL, and IEEE Explore, were used to conduct a thorough search.

The inclusion was open to 13 papers that were released between 2017 and 2021. Three topics emerged: the most popular algorithm for ischemic heart disease prediction, the reliability of ischemic heart disease prediction algorithms, and clinical outcomes to raise the standard of treatment. Both supervised and unsupervised machine learning have been used in all approaches.

2. **L.S.T.M model**

The large-scale patient hospital records are not successfully employed to improve the prediction performance, and previous dynamic prediction models seldom handle multi-period data with variable intervals. Some studies use an enhanced long short-term memory (LSTM) model to examine the prediction of cardiovascular disease.

3. **S.V.M (Support Vector Machine) algo**

One of the machine learning algorithms is called the support vector machine. An algorithm for supervised learning is the support vector machine. The provided data is categorized using the support vector machine. A hyper plane is used by the method to distinguish between the various classes. Regression analysis also makes use of support vector machines. Both linear and non-linear data are classified by SVM.The SVM classifier's primary goal is to locate the hyperplane in an n-dimensional space.

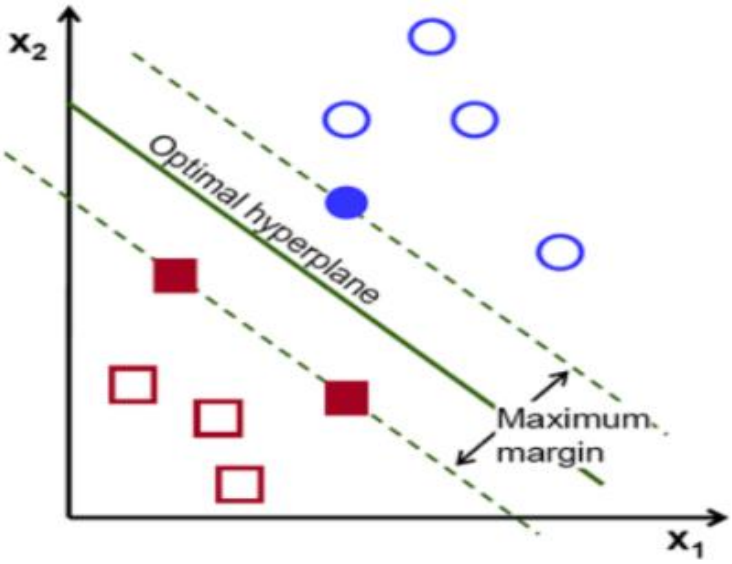

Fig 4.2 showing S.V. M

4. **Random Forest**

Machine learning algorithms are increasingly being used to forecast different illnesses. This idea is so vital and useful because machine learning algorithms are designed to think like humans. Here, the problem of improving heart disease prediction accuracy is tackled. The Cleveland heart disease dataset's non-linear tendency was taken advantage of when Random Forest was used, yielding an accuracy of 85.81%. We achieve more accuracy when utilizing the Random Forest approach to forecast cardiac illnesses with well-defined features. 303 data instances were used to train Random Forest, and 10-fold cross validation was used to verify correctness. Future lives might be spared by the suggested method for heart disease prediction.

5. **Decision Tree**

A typical data mining technique for creating classification and prediction systems based on many explanatory characteristics and creating prediction models for a target instance is the decision tree methodology. This technique divides a population into pieces that resemble branches in a tree that create a root node, internal nodes, and leaf nodes in an inverted tree. A decision tree is a non-parametric technique that can handle large, complex data sets effectively without utilizing several parametric structures. Study data can be split into training and validation data sets if the sample size is big enough.

Choose the right tree size to get the best final model by using the training data set to create a decision tree model and the validation data set.

6. **Logistic Regression**

The link between the dependent variable (target), which is categorical data with a

nominal or ordinal scale, and the independent variable (predictor), which is categorical data with an interval or ratio scale, is assessed using the predictive model known as logistic regression. To determine the link between the relevant variables, this approach may also be employed in time series modeling. An approach called logistic regression is used to forecast the likelihood of categorical dependent variables.

7. **ANN (Artificial Neural Network)**

ANN algorithm is based on a large number of basic neural units (artificial neurons), which are roughly equivalent to the observed behavior of the axons in a real brain. It is used in computer science and other study areas. Each neuronal unit is interconnected with several others, and these connections can either increase or decrease the level of activity in nearby neural units. The outline function is used to compute for each individual neuronal unit. Each link and the unit itself may have a threshold function or limiting function that requires the signal to exceed before it may reach other neurons. These systems thrive in areas where the solution or feature identification is challenging to describe in a conventional computer programme because they are self-learning and taught rather than explicitly coded.

8. **Naive Bayes Algorithm**

Data mining is the process of applying a number of approaches to find information or decision-making expertise in a database and extracting it so that it may be used for tasks like decision support, forecasting, estimate, and prediction. The healthcare sector gathers enormous volumes of data, which are regrettably not "mined" to reveal hidden information for wise decision-making. Data mining is the process of identifying relationships between variables in a database. The Decision Support in Heart Disease Prediction System (DSHDPS) established by this study makes use of the Naive Bayes data mining modeling approach. The chance of people developing heart disease may be predicted using medical profiles including age, sex, blood pressure, and blood sugar.It is implemented as an online survey application. It may be used as a teaching tool to teach nurses and medical students how to diagnose heart disease patients.

4.4 CONCLUSION

The goal of this study was to determine if patient questionnaires containing historical subjective and examination-based objective health data might be utilised to detect potential risks for heart disease. Such data may support the diagnostic value of physiological-biochemical tests clinically carried out in CVD in addition to screening. SVMs with strict feature selection were taken into account by the evaluation system. The categories of medical condition, cardiovascular health, and fitness have shown good promise in determining the risk of CVD after a number of tests, with the category of fitness demonstrating significant effectiveness.

Researchers have outlined many machine learning techniques for heart disease prediction. They developed a number of machine learning algorithms and then examined their attributes to determine which one was the best. Every algorithm has produced a distinct outcome in a variety of circumstances. Further analysis shows that the prediction model for heart illness only achieves minimal accuracy; hence, more complicated models are required to improve the accuracy of predicting early heart disease Future methodologies for highly accurate, low-cost, and simple early heart disease prediction will be proposed.The researchers stated many algorithms and the algorithms have problems also.But the best 3 with the accuracies are Decision Tree. ,Naive Bayes and LSTM.The references are given in the References section of this report.

CHAPTER 5
METHODOLOGY

Research methodology is **the specific procedures or techniques used to identify, select, process, and analyze information about a topic**.

5.1 Algorithms From Literature Survey

The Algorithms that can be used for the given problem statement with their accuracy percentage as well as the references paper is shown below:

S no.	Reference No.	Techniques /Methods used	Accuracy(%)
1	17	ANN	85.53
2	20	Naive Bayes	96.5
3	21	Decision Tree	99.2
4	22	SVM	86.6
5	6	Logistic Regression	83.70
6	9	Random Forest	86.9
7	15	LSTM	92.5
8	16	PRISMA	84.5

Fig 5.1 showing different techniques methods to be used

5.2 Discussion with the Expert

We went to the doctor in Sagar Hospital on 10th and 15th Nov,2022 to discuss how our project can be more useful to the medical field.They told us how the heart diseases are increasing day by day.They told us regarding a mobile application Medscape which helps in the very early stage of risk to heart disease.We meet their many doctors but the special guidance was given by Dr.Raisa(PG doctor).She told regarding the symptoms and many thing regarding that which is helpful for for project implementation.

Fig 5.1: After discussion pics with Dr.Raisa,PG Doctor,Sagar Hospital,Bengaluru

We have discussed the symptoms of different diseases.The symptoms of heart diseases are as follows:

- Chest Pain
- Shortness of Breath.
- Coughing or Wheezing.
- Swelling in the Legs, Ankles, or Feet.
- Narrowed Blood Vessels.
- Fatigue.
- Fast or Uneven Heartbeat (Palpitations) .

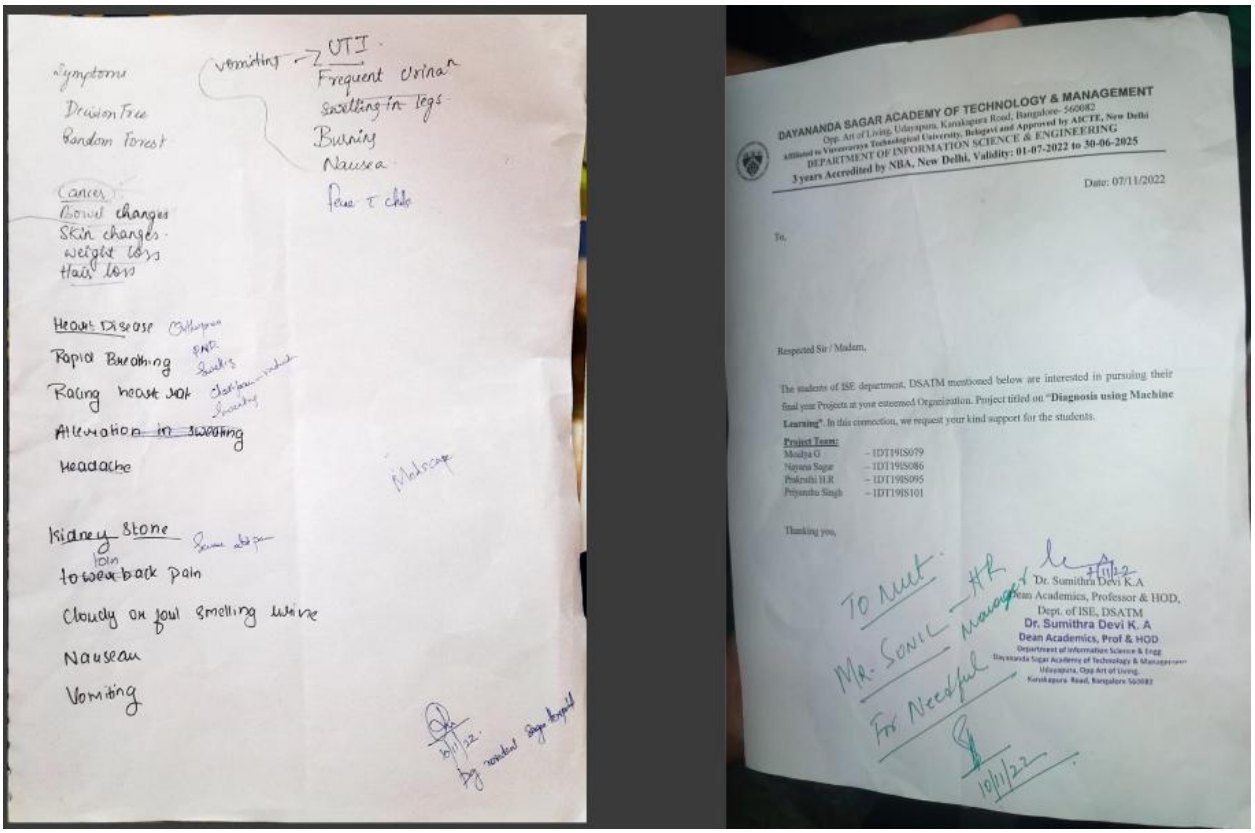

Fig 5.2 showing discussion pad used in sagar hopital and permission from college

5.3 Tools to be used(Some tools might change during the implementation)

Hardware tools:

RAM:8Gb & above

Hard disk:128 Gb & above

Software Tools:

Operating System:Windows 10 & above

Programming Language: Python 3.8 &above

Frontend Tools:HTML,CSS etc.

CHAPTER 6
CONCLUSION

We have done literature surveys,meet the experts and gone through the different methodologies.Our project will predict people with cardiovascular disease by extracting the patient medical history that leads to a fatal heart disease from a dataset that includes patients' medical history such as chest pain, sugar level, blood pressure, etc. This Heart Disease detection system will assist a patient based on his/her clinical information of them been diagnosed with a previous heart disease.The various algorithms that can be used for this project are ANN,Naive Bayes,Decision Tree etc.Python might be a great option for coding the web application.

REFERENCES

[1] A. Ishaq, S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, ''Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques,'' IEEE Access, vol. 9, pp. 39707–39716, 2021.

[2] S. Singh, and R. Zeltser, "*Cardiac Risk Stratification,"* in: *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2020.

[3] H. B. Kibria and A. Matin, ''The severity prediction of the binary and multi-class cardiovascular disease—A machine learning-based fusion approach,'' Comput. Biol. Chem., vol. 98, Jun. 2022, Art. no. 107672.

[4] NHANES: https://www.cdc.gov/nchs/nhanes/index.htm

[5] C.Y. Wang, et al., "Cardiorespiratory fitness levels among US adults 20-49 years of age: findings from the 1999-2004 National Health and Nutrition Examination Survey," *Am J Epidemiol*., vol. 171, no. 4, pp. 426-435, Feb. 2010.

[6] P.L. Tsou, and C.J. Wu, "Sex-Dimorphic Association of Plasma Fatty Acids with Cardiovascular Fitness in Young and Middle-Aged General Adults: Subsamples from NHANES 2003-2004," *Nutrients*, vol. 10, no. 10, 1558, Oct. 2018.

[7] S.S. Yoon, et al., "Trends in the Prevalence of Coronary Heart Disease in the U.S.: National Health and Nutrition Examination Survey, 2001-2012," *Am. J. Prev. Med*., vol. 51, no. 4, pp. 437-445, Oct. 2016.

[8] R. Moonesinghe, et al., "Prevalence and Cardiovascular Health Impact of Family History of Premature Heart Disease in the United States: Analysis of the National Health and Nutrition Examination Survey, 2007-2014," *J. Am. Heart Assoc*., vol. 8, no. 14, e012364, July 2019.

[9] K. Jindai, et al., "Multimorbidity and Functional Limitations Among Adults 65 or Older, NHANES 2005–2012," *Prev. Chronic Dis*., vol. 13, 160174, Nov. 2016.

[10] Thirukrishna, J.T., Krishna, S.R.S., Shashank, P. Survey on Diagnosing CORONA VIRUS from Radiography Chest X-ray Images Using Convolutional Neural Networks. Wireless Pers Commun (2022) (Springer).

[11] A. Koyanagi, et al., "Correlates of physical activity among community-dwelling adults aged 50 or over in six low- and middle- income countries," *PLoS ONE*, vol. 12, no. 10, e0186992, Oct. 2017.

[12] W.-H. Weng, "Machine Learning for Clinical Predictive Analytics," in: *Leveraging Data Science for Global Health*. L. A. Celi et al. (eds.), 2020, ch. 12.

[13] JT Thirukrishna, Aishwarya MV, Mansi Singh, Mounisha B, and Naksha Kaveri. "Efficient data Transmission in WSN using wearable sensors for Healthcare Monitoring." International Journal of Advance Research And Innovative Ideas In Education 7, no. 2 (2021) : 446-457.

[14] H. Sanz, et al., "SVM-RFE: selection and visualization of the most relevant features through non-linear kernels," *BMC Bioinformatics*, vol. 19, 432, Nov. 2018.

[15] A. Dinh, et al. A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. BMC Med Inform Decis Mak 19, 211, 2019. https://doi.org/10.1186/s12911-019- 0918-5

[16] Thirukrishna, J.T., Krishna, S.R.S., Shashank, P. Survey on Diagnosing CORONA VIRUS from Radiography Chest X-ray Images Using Convolutional Neural Networks. Wireless Pers Commun (2022) (Springer).

[17] . J. K. Kim and S. Kang, "Neural Network-Based Coronary Heart Disease Risk Prediction Using Feature Correlation Analysis," Journal of Healthcare Engineering, vol. 2017, 2017. 23. J. PING LI,.A. U. H.

[18] S. U. D. J. K. A. KHAN and A. SABOOR, "Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare," IEEE Access, vol. 8, 19 June 2020.

[19] M. A. Jabbar,Shirina Samreen ,"Heart disease prediction system based on hidden naive bayes classifier",IEEE,2020

[20] Mai Shouman,Tim Turner,Rob Stocker,School of Engineering and Information Technology,University of New South Wales at Australian Defence Force Academy Canberra ACT 2600,"Using Decision Tree for Diagnosing Heart Disease Patients",2021

[21] T Mythili,Dev Mukherji,Nikita Padalia,Abhiram Naidu,International Journal of Computer Applications,"A heart disease prediction model SVM-Decision Trees-Logistic Regression(SDL)",2013

[22] G. Magesh and P. Swarnalatha, ''Optimal feature selection through a cluster-based DT learning (CDTL) in heart disease prediction,'' Evol. Intell., vol. 14, no. 2, pp. 583–593, Jun. 2021.

[23] A. Pandya, et al., "A comparative assessment of non-laboratory- based versus commonly used laboratory-based cardiovascular disease risk scores in the NHANES III population," *PLoS One*, vol. 6, no. 5, pp. e20416, May 2011.

[24] WHO, https://www.who.int/health-topics/cardiovascular- diseases#tab=tab_1