Project Synopsis

on

**PENTACKLES**

Submitted as a part of course curriculum for

**Bachelor of Technology**

in

**Computer Science**



**Submitted by**

Aakansha Tyagi (2000290120001)

Aakriti Gupta (2000290120002)

Ayush Jaiswal (2000290120052)

Ayush Tyagi (2000290120056)

**Under the Supervision of**

Mr. Vivek Kumar Sharma

Professor

**KIET Group of Institutions, Ghaziabad**

**Department of Computer Science**

**Dr. A.P.J. Abdul Kalam Technical University**

**2022-2023**

**INDEX**

|  |  |  |
| --- | --- | --- |
| **S NO:** | **ESSENTIALS** | **PAGE NO:** |
| **1** | TITLE PAGE |  |
| **2** | DECLARATION |  |
| **3** | CERTIFICATE |  |
| **4** | ACKNOWLEDGEMENT |  |
| **5** | ABSTRACT |  |
| **6** | CHAPTER 1 INTRODUCTION |  |
| **7** | 1.1. Introduction |  |
| **8** | 1.2 Problem Statement |  |
| **9** | 1.2. Objective |  |
| **10** | 1.3. Scope |  |
| **11** | CHAPTER 2 LITERATURE REVIEW |  |
| **12** | CHAPTER 3 PROPOSED METHODOLOGY |  |
| **13** | 3.1 Flowchart |  |
| **14** | CHAPTER 4 TECHNOLOGY USED |  |
| **15** | CHAPTER 5 CONCLUSION |  |
| **16** | REFERENCES |  |
|  |  |  |

**DECLARATION**

We hereby declare that this submission is our work and that to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgement has been made in the text.

Signature of Student

Name: Aakansha Tyagi

Roll No.:2000290120001

Date:

Signature of Student

Name: Aakriti Gupta

Roll No.:2000290120002

Date:

Signature of Student

Name: Ayush Jaiswal

Roll No.:2000290120052

Date:

Signature of Student

Name: Ayush Tyagi

Roll No.:2000290120056

Date:

**CERTIFICATE**

This is to certify that Project Report entitled “**PENTACKLES**” which is submitted by **Aakansha Tyagi, Aakriti Gupta, Ayush Jaiswal, Ayush Tyagi,** in partial fulfilment of the requirement for the award of degree B. Tech. in Department of Computer Science of Dr A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

**Date: Supervisor Signature**

Mr. Vivek Kumar Sharma

(Professor)

**ACKNOWLEDGEMENT**

It gives us a great sense of pleasure to present the synopsis of the B.Tech Mini Project undertaken during B.Tech. Third Year. We owe a special debt of gratitude to **Mr. Vivek Kumar Sharma** (Professor), Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his/her cognizant efforts that our endeavours have seen the light of the day.

We also take the opportunity to acknowledge the contribution of **DR. AJAY KUMAR SHRIVASTAVA** (Head of the Department of Computer Science), KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

Last but not the least, we acknowledge our friends for their contribution to the completion of the project.

Signature of Student

Name: Aakansha Tyagi

Roll No.:2000290120001

Date:

Signature of Student

Name: Aakriti Gupta

Roll No.:2000290120002

Date:

Signature of Student

Name: Ayush Jaiswal

Roll No.:2000290120052

Date:

Signature of Student

Name: Ayush Tyagi

Roll No.:2000290120056

Date:

**ABSTRACT**

Machine learning and Artificial Intelligence are playing a huge role in today’s world. From self-driving car, medical fields, we can find them everywhere. The medical industry generates a huge amount of patient data which can be processed in a lot of ways. So, with the help of machine learning, we have created a Prediction System that can detect more than one disease at a time. Many of the existing systems can predict only one disease at a time and that too with lower accuracy. Lower accuracy can seriously put a patient’s health in danger. We have considered three diseases for now that are Heart, Liver, and Diabetes and in the future, many more diseases can be added. The user has to enter various parameters of the disease and the system would display the output whether he/she has the disease or not. This project can help a lot of people as one can monitor the persons’ condition and take the necessary precautions thus increasing the life expectancy.

**CHAPTER 1- INTRODUCTION**

**1.1 Introduction**

The system we have proposed is user friendly to get help and advice on health issues immediately through the online healthcare system. Now a days, with the help of the statistics and posterior distribution the problems are swiftly and easily. As the Bayesian statistics has a great success rate in the field of economic, social science and a few other fields just like that, in medical fields, people have solved various medical problems that are tiresome to be settled in classic statistics by classification and can be solved easily.The classification rules which help in solving the prediction of disease are generated by the samples trained by themselves and help in solving the problem easily.

It is approximated that greater than 70% of people in India are prone to various body diseases like viral, flu, cough, cold etc. in intervals of 2 months. As many people don’t understand that the general body diseases could be symptoms of something more harmful, 25% of this population dies or gets some serious medical problem because of ignoring the early general body symptoms and this is a very serious condition that we are facing and the problem can be proven to be a very dangerous situation for the population and can be alarming if the people will continue ignoring these diseases. Hence identifying or predicting the disease at the very basic stage is very important to avoid any unwanted problems and deaths. The systems which are available now a days are the systems that are either dedicated to a particular disease or are in development or the research for solving the algorithms related to the problem when it comes to generalized disease.

The main motive of the proposed system is the prediction of the commonly occurring diseases in the early phase as when they are not checked or examined they can turn into a disease more dangerous disease and can even cause death. The system applies data mining techniques , decision tree algorithms, Naive Bayes algorithm and Random Forest algorithm. This system will predict the most possible disease based on the given symptoms by the user and precautionary measures required to avoid the aggression of disease, it will also help doctors to analyze the patterns of diseases in the society. This project is dedicated to the Disease prediction System that will have data mining techniques for the basic stages of the dataset and the main model will be trained using the Machine Learning (ML) algorithms and will help in the prediction of general diseases.

### **Data Analysis and Data Mining**

The Data Mining is a process in which raw data is prepared and structured from the unstructured data as to take meaningful information from the data which can be used in the project. Task of making data organized and reflective about data is to way to get what this information does the data contains in it and what it does not have in it. There are so many different types of methods in which the people can make use of data analysis. It is simply very easy to use data during the analysis phase and get to some certain conclusions or some agendas. The analysis of data is a process of inspecting, cleaning, transforming, and modelling data with the objective of highlighting useful information, suggesting conclusions, and supporting decision making which are helpful to the user. Data analysis has multiple facets and approaches, encompassing diverse techniques under an array of names, in different business, science, and social science domains.

### **Machine Learning Algorithms**

KNN Algorithm

Random Forest Algorithm

XG Boost Algorithm

**1.2** **Problem Statement**

Many of the existing machine learning models for health care analysis are concentrating on one disease per analysis. For example first is for liver analysis, one for cancer analysis, one for lung diseases like that. If a user wants to predict more than one disease, he/she has to go through different sites. There is no common system where one analysis can perform more than one disease prediction. Some of the models have lower accuracy which can seriously affect patients’ health. When an organization wants to analyse their patient’s health reports, they have to deploy many models which in turn increases the cost as well as time Some of the existing systems consider very few parameters which can yield false results.

**1.3 Objective**

In multiple disease prediction, it is possible to predict more than one disease at a time. So the user doesn’t need to traverse different sites in order to predict the diseases. We are taking three diseases that are Liver, Diabetes, and Heart. As all the three diseases are correlated to each other. To implement multiple disease analyses we are going to use machine learning algorithms and Django. When the user is accessing this API, the user has to send the parameters of the disease along with the disease name. Django will invoke the corresponding model and returns the status of the patient.

**1.4 FUTURE SCOPE**

\*In the future we can add more diseases in the existing API

**\***We can try to improve the accuracy of prediction in order to decrease the mortality rate.

**CHAPTER 2- LITERATURE REVIEW**

**\*REVIEW 1**

* **Title: A Hybrid Cauchy Crazy Particle Swarm Optimization Support Vector Machine Algorithm for Heart disease prediction**
* **Author:** Mandakini Priyadarshani Behera et al.
* **Journal Name:** Procedia Computer Science
* **Year of Publishing:** 2022
* **Summary-**

1. Provided an algorithm for heart and liver disease prediction, as well as heart and liver datasets. SVM, PSOSVM, CPSOSVM, and CCPSOSVM are four algorithms used to predict heart and liver disease. The performance of each method has been calculated and assessed in terms of the confusion matrix, classification accuracy, classification error rate, precision, recall, and F1 score**.**
2. The heart disease dataset was collected from the University of California, Irvine (UCI), machine learning repository. The heart dataset has 13 attributes and 270 instances, the dataset for liver disease is retrieved from the UCI machine learning repository and the dataset has 10 attributes and 583 instances**.**
3. The study concludes from rigorous experimental investigation that designed CCPSOSVM gives excellent classification results, with the highest classification rate and lowest error rate for heart and liver sickness prediction.

**\*REVIEW 2**

* **Title: Machine Learning Model Prediction of Mortality in Patients With and Without Heart Failure**
* **Authors:** Se Yong Jang et al**.**
* **Journal Name:** JACC: Advances
* **Year of Publishing:** September 2023
* **Summary:**

1. Predicting mortality risk is vital for tailoring medical care to patients, allowing for appropriate monitoring, therapy referral, and end-of-life counseling. Existing risk prediction models tend to be disease-specific, limiting their applicability across different medical conditions. The MARKER-HF model aims to overcome this limitation by providing a reliable indicator of mortality risk across various diseases.
2. The result of this study shows that the average age of the 41,749 patients was 65, with 56.2% being male**.**
3. MARKER-HF accurately predicted mortality in subgroups with and without cardiovascular disease, as well as in patients with acute coronary syndrome, atrial fibrillation, chronic obstructive pulmonary disease, chronic kidney disease, diabetes mellitus, or hypertension. . The study concludes that MARKER-HF predicts mortality for patients with HF as well as for patients suffering from a variety of diseases.

**\*REVIEW 3**

* **Title: A Novel Machine Learning Model with Stacking Ensemble Learner for Predicting Emergency Readmission of Heart-Disease Patients**
* **Author:** Alireza Ghasemieh, Alston Lloyed, Parsa Bahrami, Pooyan Vajar, Rasha Kashef
* **Journal Name:** Decision Analytics Journal
* **Year of published:**2023
* **Summary-**

1. Cardiovascular diseases are a leading cause of mortality worldwide, emphasizing the need for early detection and intervention. Existing machine learning approaches for heart disease detection have limitations, resulting in misdiagnoses and overcrowded medical facilities. This study introduces a novel model for identifying patients at risk of emergency readmission, utilizing behavior-based features to create a new class label for emergency readmission**.**
2. The proposed model employs a robust Stacking Ensemble Learner (SEL) using ensemble learning, with XGBoost as the meta-learner. The model predicts whether a patient with heart problems requires emergency admission after an initial admission.
3. The study acknowledges the potential limitation of the dataset size, which might affect the generalizability of the model.

**\*REVIEW 4**

* **Title-Using a machine learning-based risk prediction model to analyze the coronary artery calcification score and predict coronary heart disease and risk assessment**
* **Author -**Yue Huang a,1 , YingBo Ren a,1 , Hai Yang a , YiJie Ding b , Yan Liu a , YunChun Yang a , AnQiong Mao a , Tan Yang d , YingZi Wang c , Feng Xiao c , QiZhou He e, Ying Zhang a
* **Journal Name-**Computers in Biology and Medicine
* **Year of publishing-**2022
* **Summary-**

1. Cardiovascular disease, particularly coronary heart disease (CHD), is a leading cause of death worldwide. The study aims to evaluate the use of ML models to predict CHD risk more accurately.
2. Conditions such as hypertension, diabetes, fatty liver, osteoporosis, and hyperlipidemia were considered. Various ML models were selected for prediction, including Random Forest (RF), k-nearest neighbor(KNN), support vector machines (SVM), Kernel Ridge Regression (KRR), and radial basis function neural networks (RBFNN). Performance evaluation of the models included metrics such as accuracy, sensitivity, specificity, and the Matthews correlation coefficient (MCC).
3. RF was found to have the best predictive performance in terms of accuracy, sensitivity, specificity, MCC, and AUC compared to other ML models**.**

**\*REVIEW 5**

* **Title-Development of smart cardiovascular measurement system using feature selection and machine learning models for prediction of sleep deprivation, cold hands and feet, and Shanghuo syndrome**
* **Author Name-** Chun-Ling Lin a, Chin-kun Tseng b, Chien-Jen Wang b, Shu-Hung Chao b, Yuh-Shyan Hwang b, Lih-Jen Kau b
* **Journal Name-** International Measurement Confederation
* **Year of publishing-**2023
* **Summary-**

1. It introduces the concept of a smart cardiovascular measurement system that utilizes ECG and PPG to evaluate sleep deprivation, cold hands and feet, and Shanghuo syndrome (from Traditional Chinese Medicine) and discusses the goals of the study
2. The study involved 83 healthy adults, including 26 males and 57 females. Features were extracted from both ECG and PPG signals. A total of 38 features were selected, including 19 from ECG and 19 from PPG. Various machine learning methods were explored, including ensemble learning, kernel classification, k-nearest neighbor, support vector machine, naive Bayes, neural network, and decision tree classifiers. Bayesian optimization was used to fine-tune model selection and hyperparameter values. Model performance was measured using accuracy, computed using true positives, true negatives, false positives, and false negatives.
3. The study acknowledges limitations, including a relatively small sample size and the potential influence of additional factors not considered.

**\*REVIEW 6**

* **Title: Machine Learning Approach for Risk Factors Analysis and Survival Prediction of Heart Failure Patients**
* **Authors:** Md. Mamun Ali a , Vian S. Al-Doori b , Nubogh Mirzah c , Asifa Afsari Hemu d , Imran Mahmud a , Sami Azam e , Kusay Faisal Al-tabatabaie f , Kawsar Ahmed g,h,∗ , Francis M. Bui g , Mohammad Ali Moni
* **Journal Name:** Healthcare Analytics
* **Published Year:**2023
* **Summary:**

1. Cardiovascular diseases, including HF, are a major global health concern. This study focuses on HF, a condition where the heart struggles to pump enough blood. The research aims to analyze risk factors and predict survival in HF patients using machine learning methods.
2. The study employs a dataset with 299 instances, including 96 deaths and 203 survivors. Data preprocessing involves handling missing values, feature engineering, and balancing the dataset.
3. Five ML algorithms are applied, including Decision Tree, Random Forest, XGBoost, and Gradient Boosting, to train and test the models. Performance metrics (accuracy, precision, recall, F-measure, log loss) are employed for evaluation. RF shows maximum accuracy in this model**.**

**\*REVIEW 7**

* **Title: Machine Learning-Based Predictive Models for Heart Disease Diagnosis**
* **Author Name:** Stojanov, Lazarova, Veljkova, Rubartelli, Giacomini
* **Journal Name:** King Saud University
* **Year of Publishing:** 2023
* **Summary:**

1. This study explores the use of machine learning techniques to predict heart disease based on various health-related attributes. Feature selection is employed to identify relevant features, and ten different machine learning algorithms are analyzed. The Decision Tree algorithm stands out with the highest accuracy .
2. This study leverages logistic regression as a core model to predict heart failure against chronic-ischemic heart disease in the elderly population. By analyzing biochemical variables, a subset of variables is identified as excellent for discriminating between heart failure and chronic-ischemic heart disease.
3. The study highlights the predictive potential of individual biochemical parameters. While the study faced limitations due to a small sample size, it offers the groundwork for the development of precise diagnostic software, which can be enhanced with more patient data

**\*REVIEW 8**

* **Title: Heart Diseases Prediction based on Stacking Classifiers Model**
* **Author:** Subasish Mohapatra, Sushree Maneesha, Prashanta Kumar Patra, Subhadarshini Mohanty
* **Journal Name:** International Conference on Machine Learning and Data Engineering
* **Year of Publishing:** 2023
* **Summary:**

1. The introduction highlights the importance of machine learning in various domains, including healthcare. It emphasizes the need for efficient analysis of the growing volume of medical data generated by digital systems.
2. Cardiovascular diseases are a significant global health concern, and early detection is crucial for effective treatment. Traditional diagnostic methods are slow and costly, motivating the adoption of machine learning for faster and more accurate diagnoses.
3. The methods section outlines the workflow, data preprocessing steps and the stacking approach used for model building. The stacking involves two levels: base learners and meta-learners. The goal is to harness the strengths of diverse classifiers to improve prediction accuracy.

**\*REVIEW 9**

* **Title: Automating Cardiovascular Disease Prediction Using Machine Learning and EMR Data**
* **Authors:** Qi Li, Alina Campan, Ai Ren, Wael E. Eid
* **Journal:** International Journal of Medical Informatics
* **Year of Publishing:** 2022
* **Summary:**

1. Cardiovascular disease is a significant health concern globally, with varying prevalence and mortality rates in different regions. Accurate risk assessment is crucial for effective prevention and intervention strategies. Current risk assessment tools like the PCE Risk Calculator may not provide precise risk estimates for all populations. This study aims to develop an automated CVD risk calculator tailored to a specific population using machine learning and EMR data.
2. The authors collected EMR data from over 100,000 patients from a regional healthcare system, spanning from January 1, 2009, to April 30, 2020. Machine learning techniques were applied to these datasets, considering both cross-sectional (CS) features and a combination of CS and longitudinal (LT) features derived from vital statistics and laboratory values. The study evaluates the performance of various machine learning models and compares them to the PCE Risk Calculator.
3. All machine learning models tested outperformed the PCE Risk Calculator in predicting CVD risk. The random forest machine learning technique applied to the combination of CS and LT features achieved the highest accuracy.

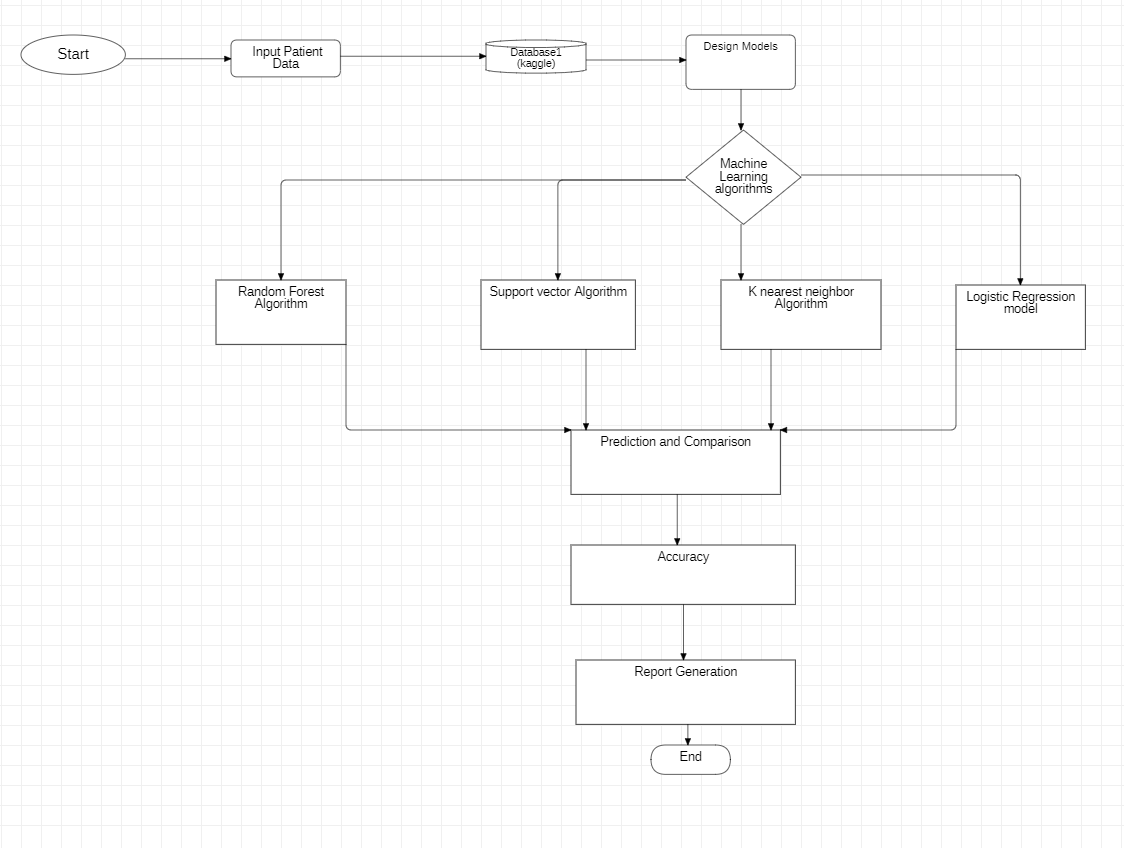
**\*REVIEW 10**

* **Title**: **Machine Learning-Based Approach to the Diagnosis of Cardiovascular Disease Using a Combined Dataset**
* **Author Names**: Khandaker Mohammad Mohi Uddin, Rokaiya Ripa, Nilufar Yeasmin, Nitish Biswas, Samrat Kumar Dey
* **Journal Name**: Intelligence-Based Medicine
* **Year of Publishing**: 2023
* **Summary:**

1. Heart disease is one of the most serious ailments, killing the majority of its victims. Heart disease is very difficult to diagnose medically. Early identification of heart disease will reduce the risk of mortality. Predicting heart illness has grown to be one of the most challenging medical tasks in recent years due to the prevalence of cardiac issues.
2. In this work, the existence of cardiac anomalies is detected using Machine Learning (ML) approaches. Several machine learning (ML) algorithm techniques, including Decision Tree (DT), Ada-Boost Classifier (AB), Extra Trees Classifier (ET), Support Vector Machine (SVM), Gradient boost, MLP, extreme gradient boost (XGB), Random Forest (RF), KNN, and LR, are used in the proposed method to predict the likelihood of heart disease and classify the risk level of the patient.
3. The testing findings demonstrate that the **Decision Tree** method has the best accuracy, when compared to other machine learning techniques.

**CHAPTER 3-PROPOSED METHODOLOGY**

**3.1Flowchart**

****

The first step is to the dataset for heart disease, diabetes disease and liver

disease we have imported the UCI dataset, PIMA dataset and Indian liver dataset respectively. Once we have imported the dataset then visualization of each imputed data takes place. After visualization pre-processing of data takes place where we check for outliers, missing values and also scale the dataset then on the updated dataset we split the data into training and testing. Next is on the training dataset we had applied K-NN and random forest algorithm and applied knowledge on the classified algorithm using testing dataset. Then we build a pickle file for all the disease and then integrated the pickle file with the Django framework for the output of the model on the webpage.

**CHAPTER 4- TECHNOLOGY USED**

**FRONTEND**

* Hypertext Markup language(HTML)
* Cascading Style Sheet(CSS)
* JavaScript(JS)
* Bootstrap

**BACKEND**

* FLASK
* Python(Numpy, Pandas, Seaborn, Matplotlib)

**Machine Learning**

* Support Vector Machine(SVM)
* Random Forest
* Logistic Regression
* K Nearest Neighbors(KNN)

**Heart disease**

Urban areas record between 400 or 500 cases in every 100,000 people, while rural populations record 100 cases per 100,000 people.

****

**CHAPTER 6- CONCLUSION**

* The main objective of this project was to create a system that would predict more than one disease and do so with high accuracy. Because of this project the user doesn’t need to traverse different websites which saves time as well. Diseases if predicted early can increase your life expectancy as well as save you from financial troubles. For this purpose, we have used various machine learning algorithms like Support Vector Machine (SVM) , Random Forest , Logistic Regression , K Nearest Neighbors (KNN) achieve maximum accuracy.

**REFERENCES**

[1] A. S. D. M. ,. S. Mandakini Priyadarshani Behera, "A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with," Procedia Computer Science, 2022.

[2] M. SeYongJang, "Machine Learning Model Prediction of Mortality in Patients With and Without Heart Failure," JACC: Advances, vol. 2, 2023.

[3] A. L. P. B. P. V. R. K. Alireza Ghasemieh, "A Novel Machine Learning Model with Stacking Ensemble Learner for Predicting Emergency Readmission of Heart-Disease Patients," Decision Analytics Journal, 2023.

[4] 1. ,. Y. R. a. ,. H. Y. a. ,. Y. D. b. ,. Y. L. a. ,. Y. Y. a. ,. A. M. a. ,. T. Y. d. ,. Y. W. c. ,. F. X. c. ,. Q. H. e. Y. Z. a. Yue Huang a, "Using a machine learning-based risk prediction model to analyze the coronary artery calcification score and predict coronary heart disease and risk assessment," Computers in Biology and Medicine, 2022.

[5] C.-k. T. b. C.-J. W. b. S.-H. C. b. Y.-S. H. b. L.-J. K. b. Chun-Ling Lin a, "Title-Development of smart cardiovascular measurement system using feature selection and machine learning models for prediction of sleep deprivation, cold hands and feet, and Shanghuo syndrome," International Measurement Confederation , 2023.

[6] V. S. A.-D. b. ,. N. M. c. ,. A. A. H. d. ,. I. M. a. ,. S. A. e. ,. K. F. A.-t. f. ,. K. A. g. ,. F. M. B. g. ,. M. A. M. Md. Mamun Ali a, "Machine Learning Approach for Risk Factors Analysis and Survival Prediction of Heart Failure Patients," Healthcare Analytics, 2023.

[7] L. V. R. G. Stojanov, ": Machine Learning-Based Predictive Models for Heart Disease Diagnosis," King Saud University, 2023.

[8] S. M. P. K. P. S. M. Subasish Mohapatra, "Heart Diseases Prediction based on Stacking Classifiers Model," International Conference on Machine Learning and Data Engineering, 2023.

[9] A. C. A. R. W. E. E. Qi Li, "Automating Cardiovascular Disease Prediction Using Machine Learning and EMR Data," International Journal of Medical Informatics, 2022.

[10] R. R. N. Y. N. B. S. K. D. Khandaker Mohammad Mohi Uddin, "Machine Learning-Based Approach to the Diagnosis of Cardiovascular Disease Using a Combined Dataset," Intelligence-Based Medicine, 2023.