**HEART DISEASE DETECTION USING MACHINE LEARNING**

**A PROJECT REPORT**

***Submitted by***

**ARUNKUMAR R [513319104006]**

**NITHISH KUMAR B [513319104022]**

**PANNEERSELVAM K [513319104025]**

**SURYA A [513319104041]**

***in partial fulfillment for the award of the degree***

***of***

**BACHELOR OF ENGINEERING**

***in***

**COMPUTER SCIENCE AND ENGINEERING**

**UNIVERSITY COLLEGE OF ENGINEERING ARNI**

**ANNA UNIVERSITY:: CHENNAI 600 025**

**MAY 2023**

**ANNA UNIVERSITY: CHENNAI 600 025**

**BONAFIDE CERTIFICATE**

Certified that this project report **“HEART DISEASE DETECTION USING MACHINE LEARNING”** is the bonafide work of **“ARUNKUMAR R [513319104006], NITHISH KUMAR B [513319104022], PANNEERSELVAM K [513319104025], SURYA A [513319104041]”** who carried out the project work under my supervision

**Signature Signature**

Mr. D. Prabhu Mrs. G. Julin Leeya

**Head Of The Department** **Supervisor**

Department of CSE Department of CSE

University College of Engineering Arni, University College of Engineering Arni, Thatchur, Arni – 632 326. Thatchur, Arni – 632 326.

Submitted for the Project Viva – Voce Examination held on ……………..

**INTERNAL EXAMINER EXTERNAL EXAMINER**

**ACKNOWLEDGEMENT**

We take privilege to express few words of gratitude and respect to all those who helped us in completion of this project. We thank the almighty God for giving us the opportunity of doing this project work successfully. Our heartfelt thanks to our beloved parents for their invaluable support to do this project.

With pleasure ,we express our heart filled sincere thanks to our respected Dean i/c **Dr. R. ARULARASAN M.Tech.,Ph.D.,** for giving moral support and has constant association to successful of our project. It’s our privilege to express our sincere thanks to the Faculty i/c, Dept of CSE , **Mr. D. PRABHU M.E., M.B.A.,** for his encouragement.

We also thank our project supervisor **Mrs.G.JULIN LEEYA M.TECH.,** who helped us a lot to do the project effectively and has given us a valuable guidance in completing this project successfully.

We thank our **Mr.G.DHANASEKAR M.E.,** Project Coordinator for his support and valuable suggestions during our project work.

We also extend our thanks to every member of the faculty who provided valuable academic guidance, and their co-operations to do this successful project. Finally we would like to thank to our parents, relatives and friends for their encouragement and enthusiastic Co-operation.

**ABSTRACT**

Heart disease is one of the major causes of life complicacies and subsequently leading to death. The heart disease diagnosis and treatment are very complex, especially in the developing countries, due to the rare availability of efficient diagnostic tools and shortage of medical professionals and other resources which affect proper prediction and treatment of patients. Inadequate preventive measures, lack of experienced or unskilled medical professionals in the field are the leading contributing factors... We develop a heart disease predict system that can assist medical professionals in predicting heart disease status based on the clinical data of patients. Our approaches include three steps. Firstly, we select 13 important clinical features, i.e., age, sex, chest pain type, trestbps, cholesterol, fasting blood sugar, resting ecg, max heart rate, exercise induced angina, old peak, slope, number of vessels colored, and Thal. We shall use few algortihms such as Logistic Regression, Decision Tree, Random Forest, KNN, SVM, Stochastic gradient descent, Adaboost, Xgb.

**TABLE OF CONTENT**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
|  | **ABSTRACT** | iii |
|  | **LIST OF FIGURES** | viii |
|  | **LIST OF ABBREVIATIONS** | x |
| **1.** | **INTRODUCTION**   * 1. GENERAL OVERVIEW OF DATA SCIENCE   2. DATA SCIENCE   3. ARTIFICIAL INTELLIGENCE   1.3.1 LEARNING PROCESSES  1.3.2 REASONING PROCESSES  1.3.3 SELF – CORRECTION PROCESS  1.4 PROBLEM STATEMENT | 1  3  3  4  5  6  6  7 |
| **2.** | **LITERATURE SURVEY** | 8 |
| **3.** | **METHODOLOGY**  3.1 EXISTING SYSTEM  3.2 PROPOSED SYSTEM  3.2.1 COLLECTION OF DATASET  3.2.2 PREPROCESSING DATA  3.2.3 SELECTION OF ATTRIBUTES  3.2.4 BALANCING OF DATA  3.2.5 PREDICTION OF DISEASE | 12  12  12  13  15  15  16  17 |
| **4.** | **WORKING OF SYSTEM**  4.1 SYSTEM ARCHITECTURE  4.2 MACHINE LEARNING  4.2.1 SUPERVISED LEARNING  4.2.2 UNSUPERVISED LEARNING  4.2.3 REINFORCEMENT LEARNING  4.3 ALGORITHMS  4.3.1 SUPPORT VECTOR MACHINE  4.3.1.1 CONCEPTS IN SVM  4.3.1.2 TYPES OF SVM  4.3.1.3 ADVANTAGES  4.3.1.4 DISADVANTAGES  4.3.2 LOGISTIC REGRESSION ALGORITHM  4.3.2.1 WORKING  4.3.2.2 ADVANTAGES  4.3.2.3 DISADVANTAGES  4.3.3 ADABOOST ALGORITHM  4.3.3.1 ALGORITHMS  4.3.3.2 ADVANTAGES  4.3.3.3 DISADVANTAGES  4.3.4 XGBOOST ALGORITHM  4.3.4.1 WORKING  4.3.4.2 ADVANTAGES  4.3.4.3 DISADVANTAGES  4.3.5 DECISION TREE  4.3.5.1 SELECTION MEASURES  4.3.5.2 NOTABLE TYPES OF DECISION TREE  4.3.5.3 WORKING  4.3.5.4 ALGORITHM  4.3.5.6 ADVANTAGES  4.3.5.7 DISADVANTAGES  4.3.6 RANDOM FOREST ALGORITHM  4.3.6.1 ASSUMPTIONS  4.3.6.2 ALGORITHM  4.3.6.3 ADVANTAGES  4.3.6.4 DISADVANTAGES  4.3.7 STOCHASTIC GRADIENT DESCENT  ALGORITHM  4.3.7.1 MOMENTUM  4.3.7.2 ADVANTAGES  4.3.7.3 DIADVANTAGES  4.3.8 KNN ALGORITHM  4.3.8.1 ALGORITHM  4.3.8.2 VALUE FOR K  4.3.8.3 WORKING  4.3.8.4 ADVANTAGES  4.3.8.5 DISADVANTAGES | 19  19  19  19  21  22  22  22  22  23  23  24  24  25  27  27  28  28  29  29  29  31  31  32  32  34  34  35  35  36  36  36  38  38  38  39  39  41  41  41  41  42  42  43  43  43 |
| **5** | **EXPERIMENTAL ANALYSIS**  5.1 SYSTEM ARCHITECTURE  5.1.1 HARDWARE REQUIREMENTS  5.2.2 SOFTWARE REQUIREMENTS  5.2 DATASET DETAILS  5.2.1 INPUT DATASET ATTRIBUTES  5.3 PERFORMANCE ANALYSIS  5.3.1 CONFUSION MATRIX  5.3.2 CORRELATION MATRIX  5.4 PERFORMANCE EVALUATION  5.4.1 EXISTING SYSTEM  5.4.2 NEW SYSTEM | 44  44  44  44  44  45  47  47  48  49  49  50 |
| **6.** | **SYSTEM DESIGN**  6.1 GENERAL  6.2 DATA FLOW DIAGRAM  6.3 USE CASE DIAGRAM  6.4 ACTIVITY DIAGRAM  6.5 SEQUENCE DIAGRAM  6.6 CLASS DIAGRAM | 51  51  51  52  53  54  55 |
| **7.** | **CONCLUSION**  7.1 CONCLUSION  7.2 APPENDIX  7.3 SNAPSHOT | 56  56  57  62 |
| **8.** | **REFERENCE**  8.1 REFERENCE | 67  67 |

**LIST OF FIGURES**

**FIGURE NO. TITLE PAGE NO.**

1.1 Symptoms of Heart Disease 2

3.1 Machine Learning 14

3.2 Pre – Processing of Data 15

3.3 Correlation matrix 16

3.4 Balancing of Data 17

3.5 Prediction Disease 18

4.1 System Architecture 20

4.2 Sigmoid Curve 26

4.3 Hyperplane 26

5.1 Dataset Attributes 45

5.2 Confusion Matrix 47

5.3 Correlation Matrix 48

5.4 Existing system performance 49

5.5 New System Performance 50

6.1 Data Flow Diagram Level 0 51

6.2 Data Flow Diagram Level 1 52

6.3 Use Case Diagram 52

6.4 Activity Diagram 53

6.5 Sequence Diagram 54

6.6 Class Diagram 55

7.1 Home Page 62

7.2 Working Page 62

7.3 Predict Page 63

7.4 About Page 64

7.5 Contact Page 64

7.6.1 Result Page 65

7.6.2 Result Page 66

**LIST OF ABBREVIATIONS**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **ABBREVIATIONS** | **EXPANSION** |
| 1 | RBC | Red Blood Cell |
| 2 | AI | Artificial Intelligence |
| 3 | SAS | Statistical Analysis System |
| 4 | UCI | University of California |
| 5 | VPA | Virtual Payment Address |
| 6 | ML | Machine Learning |
| 7 | CP | Chest Pain |
| 8 | FBS | Fasting Blood Sugar |
| 9 | SVM | Support Vector Machine |
| 10 | KNN | K-Nearest Neighbours |
| 11 | ADA | Adaptive Boosting |
| 12 | XGB | Extreme Gradient Boosting |
| 13 | SGD | Stochastic Gradient Descent |
| 14 | GBM | Gradient Boosting Machine |
| 15 | Chol | Cholesterol |
| 16 | CA | Cardiac Arrest |
| 17 | UML | Unified Modeling Language |
| 18 | DFD | Data Flow Diagram |
| 19 | np | Numpy |
| 20 | pd | Pandas |
| 21 | plt | Matplotlib.Pyplot |