Table of Contents

Table of Figures

Table of Tables

Abstract

The Heart Attack Prediction System, a supervised classification problem which uses the Decision Tree method from machine learning, is an important tool for anticipating heart attacks. This approach tackles urgent worldwide concerns associated with cardiovascular illnesses, stressing the need of early identification in successful management and prevention. The model provides accurate predictions based on critical criteria such as age, gender, cholesterol levels, blood pressure, and lifestyle. The major objective is to overcome the limits of standard diagnostic procedures by providing healthcare practitioners with insights about high-risk populations and allowing for individualized patient care. This method not only helps to enhance cardiovascular health prevention efforts, but it also aims to minimize premature deaths, hence improving public health outcomes.

1. Introduction

1.1. Artificial Intelligence

In a rapidly changing technological world, one phenomenon stands out as a sign of innovation and limitless possibilities, Artificial Intelligence(AI). Artificial intelligence, or AI, refers to the simulation of human intelligence by software-coded algorithms (Frankenfield, 2023). Nowadays Artificial intelligence is prevalent in everything from cloud-based, enterprise applications to consumer apps and even embedded firmware (Frankenfield, 2023). AI integrates sensing, understanding, planning, action, and learning technologies to enable computers to imitate human intelligence, performing tasks such as decisionmaking, problem-solving, environment perception, object identification, and pattern recognition (Kanade, 2022). Combining these talents results in tasks like as self-driving cars and facial recognition systems (Kanade, 2022).

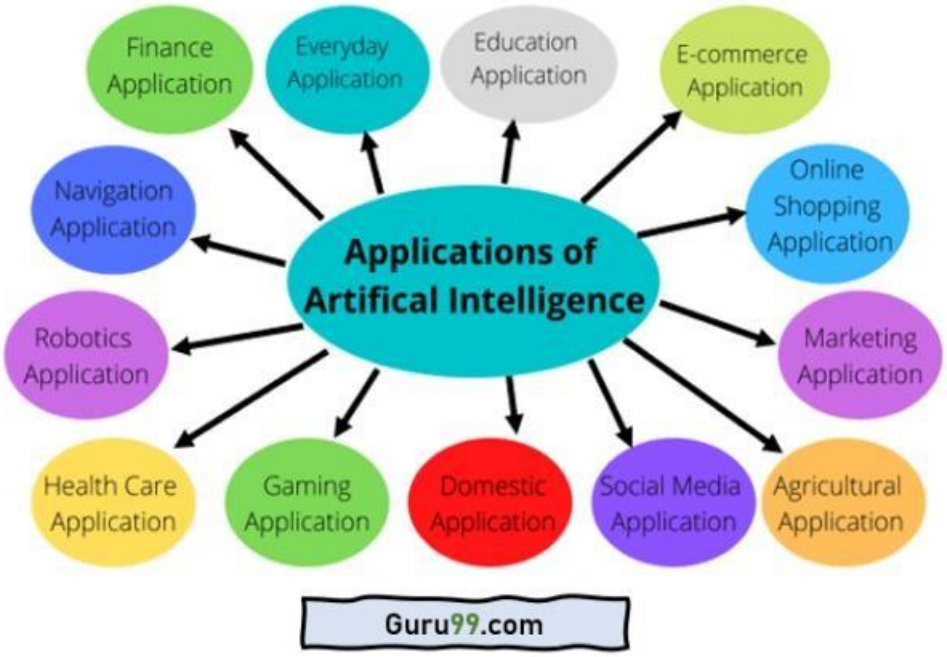

Figure 1 Application of Artificial Intelligence

AI can be divided into two categories:

AI based on Capabilities:

- **Narrow AI(Weak AI):** Narrow AI is a goal-oriented AI trained to perform a specific task. Examples of narrow AI include Apple's Siri and IBM's Watson supercomputer, use cases such as Netflix recommendations, purchase suggestions on ecommerce sites, autonomous cars etc (Kanade, 2022).
- **General AI:** General AI is an AI version that performs any intellectual task with a human-like efficiency. The objective of general AI is to design a system capable of thinking for itself just like humans do. Currently, general AI is still under research, and efforts are being made to develop machines that have enhanced cognitive capabilities (Kanade, 2022).

AI based on functionalities:

- **Reactive machines:** Reactive machines, a fundamental AI type, work without remembering prior experiences, instead concentrating entirely on present conditions to choose the best course of action. Examples include IBM's Deep Blue and Google's AlphaGo (Kanade, 2022).
- **Theory of mind:** Theory of mind refers to the type of AI that can understand human emotions and beliefs and socially interact like humans. This AI type has not yet been developed but is in contention for the future (Kanade, 2022).
- **Self-aware AI:** Self-aware AI envisions extremely intelligent computers with conscience and emotions that outperform human skills in tasks. While this notion is still a long way off, continual attempts are being made to bring it closer to reality (Kanade, 2022).

1.2. Machine Learning

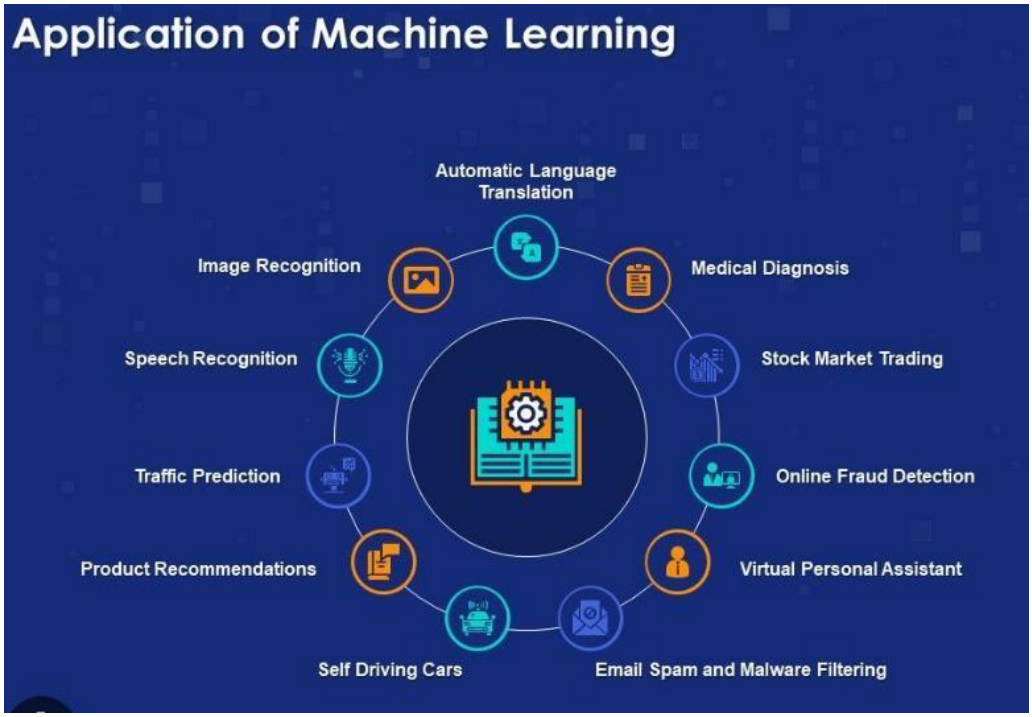

Figure 2 Application of Machine Learning

Machine Learning is categorized into two types:

- **Supervised Machine Learning:** In supervised machine learning, machines are trained on labeled datasets to predict outputs based on supplied training. The labeled dataset maps input and output parameters, allowing the machine to learn and predict using a test dataset in following rounds (Kanade, 2022). Supervised machine learning is further classified into two broad categories:
    - **Classification:** These refer to algorithms that address classification problems where the output variable is categorical (Kanade, 2022).
    - **Regression:** Regression algorithms handle regression problems where input and output variables have a linear relationship. These are known to predict continuous output variables (Kanade, 2022).

- **Unsupervised Machine Learning :** Unsupervised learning is a learning strategy that does not require supervision. In this case, the machine is trained on an unlabeled dataset and can predict the output without human intervention. An unsupervised learning algorithm attempts to group the unsorted dataset based on similarities, differences, and patterns in the input (Kanade, 2022). Unsupervised machine learning is further classified into two types:
    - **Clustering:** The clustering approach refers to putting things into groups based on factors such as similarities or differences between objects (Kanade, 2022).
    - **Association:** Association learning is the process of detecting typical relationships (Javapoint, n.d.)between variables in a big dataset. It determines the interdependence of several data items and maps the variables connected with them (Kanade, 2022).

1.3. Explanation of the chosen problem domain.

Cardiovascular diseases (CVDs) are the leading cause of mortality worldwide, claiming the lives of an estimated 17.9 million people each year (World Health Organization, 2023). Of the total CVD deaths, an estimated 7.3 million were due to coronary heart disease and 6.2 million were due to stroke. Annual CVD deaths are projected to rise to 23.3 million by 2030 (mainly from heart attacks and strokes) if current trends are allowed to continue (NCD Alliance, 2023). CVDs are a collection of heart and blood vessel illnesses that include coronary heart disease, cerebrovascular disease, rheumatic heart disease, and others.

AI in predicting heart attacks and strokes not only answers present obstacles, but also promises considerable benefits, such as improved public health outcomes and a reduction in early mortality from heart-related illnesses.

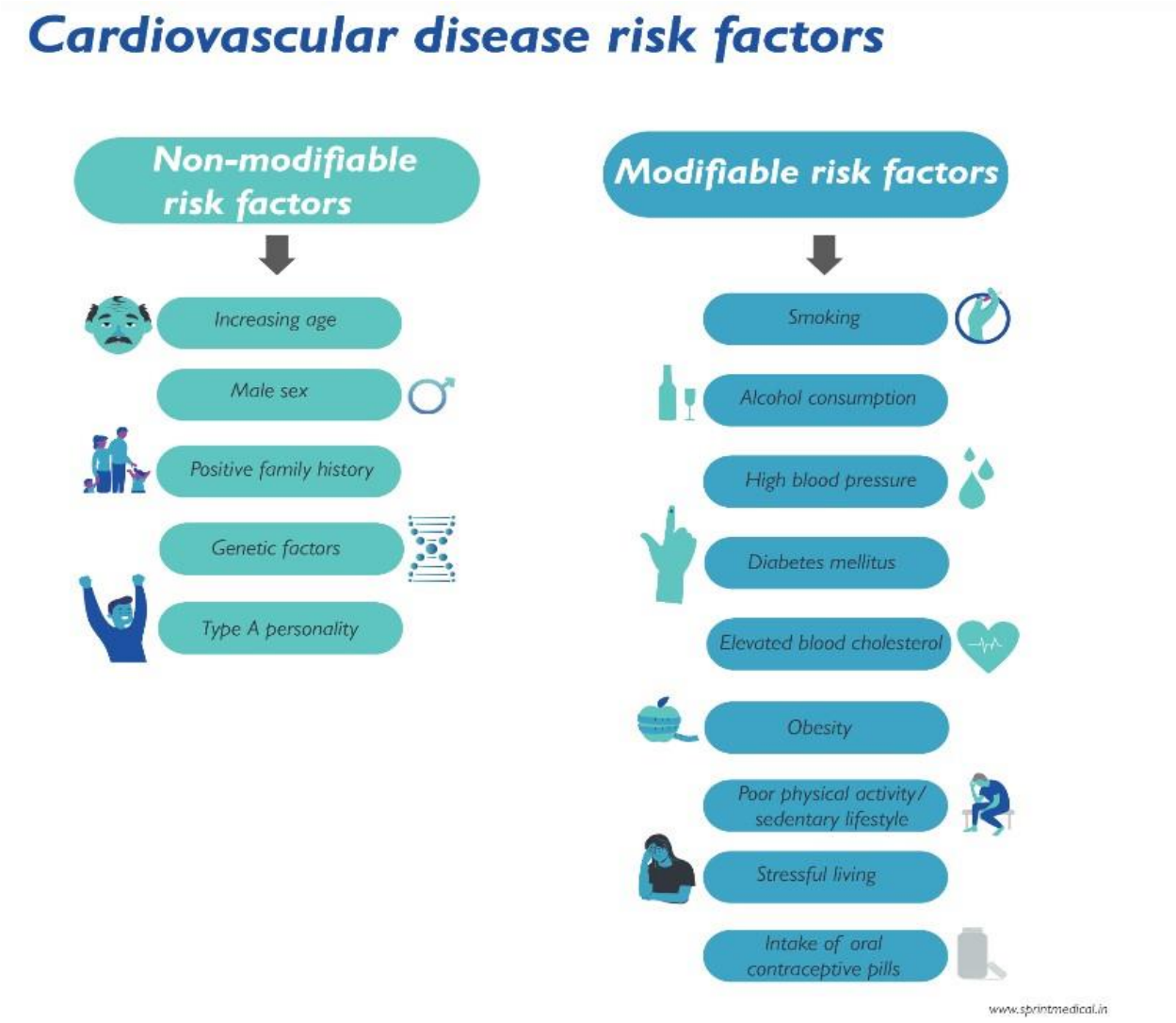

Figure 3 cardiovascular disease risk factors

2. Background

2.1. Research work done on the chosen problem domain.

Heart attacks and strokes account for more than four out of every five CVD fatalities, and one-third of these deaths occur in adults under the age of 70 (World Health Organization, 2023). The rate of premature death by CVDs has increased a lot and to prevent it requires early identification of high-risk individuals and availability to therapy (World Health Organization, 2023). The incorporation of Machine learning technology significantly decreases manual labor and increases the rate of prediction for people in risk. As an instance, machine learning can identify which people are at risk of getting heart disease or stroke and give personalized suggestions for food and exercise to avoid the condition. AI may also remotely monitor patients' health, alerting doctors to changes in patients' situations before they become critical (Avcontentteam, 2023).

An AI tool that can forecast the 10-year risk of catastrophic heart attacks, according to a study done by the University of Oxford, might alter therapy for patients who receive CT scans to check chest discomfort (University of Oxford, 2023). In the first real-world study of the AI tool, it was discovered that the AI tool improved therapy for up to 45% of patients (University of Oxford, 2023). The technique has the potential to save the lives of thousands of people suffering from chest discomfort who were not diagnosed as being at risk of a heart attack and hence did not receive adequate therapy to reduce their risk (University of Oxford, 2023). The researchers expect that this AI tool will revolutionize the way patients who are referred for chest discomfort tests are managed across the NHS (University of Oxford, 2023).

The application of AI in predicting heart attacks and strokes not only overcomes present obstacles in identifying high-risk patients, but it also has the potential for significant benefits. So, using AI in CVDs will greatly increase in public health outcomes and reducing premature deaths associated with heart-related conditions.

2.2. Review and analysis of existing work in the problem domain

2.2.1. Analysis of The Diagnostic Parameters of Heart Diseases and Prediction of Heart Attacks

Author: Gnaneswari G

Publisher: IEEE

Link: https://ieeexplore.ieee.org/document/9972211 **Overview**:

The study delves into the application of sophisticated machine learning techniques to predict heart attacks in individuals with a history of cardiac disease. Through the utilization of esteemed methods such as Support Vector Machine, Gaussian Naïve Bayes, K Nearest Neighbors, Decision Tree, Logistic Regression, Stochastic Gradient Descent, and Random Forest Classifier, this research emphasizes the significance of incorporating key statistical variables for accurate predictions. Notably, essential factors like age, sex, blood pressure, and cholesterol are included in the dataset. By subjecting a dataset of 919 patient records to feature reduction and exploratory data analysis methods, the study reveals that the Random Forest classifier demonstrates exemplary success with an impressive accuracy of 89.13%. Distinguishing itself from conventional studies that focus solely on predicting heart disease, this research makes a remarkable contribution by specifically targeting heart attack prediction in individuals who have already encountered heart disease (G, 2022).

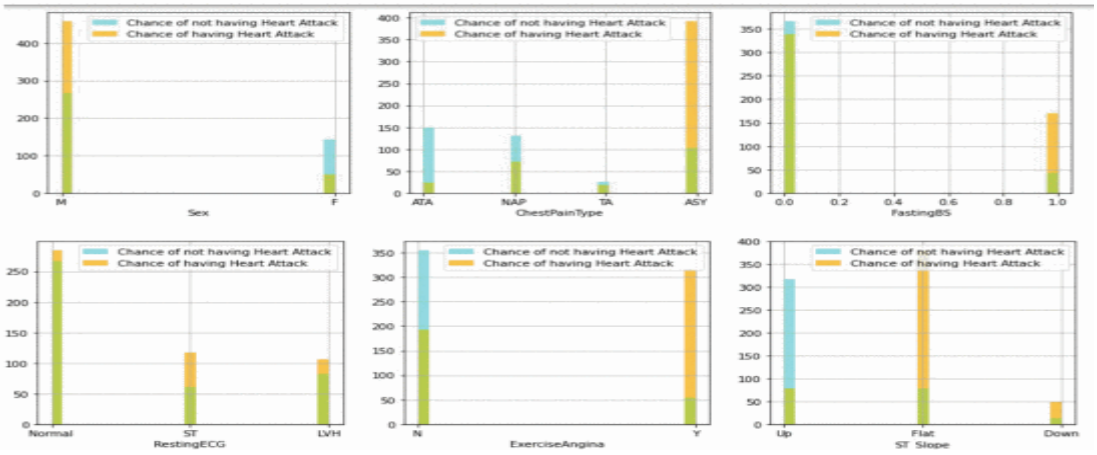

Figure 4 Data analysis of report

2.2.2. Prediction of Heart attack using Supervised Machine Learning Technique

Author: Jesslyn Audrey, Mochammad Haldi Widianto

Publisher: IEEE

Link: https://ieeexplore.ieee.org/document/9885688 **Overview**:

This study is focused on using machine learning classification models to predict heart attacks. It also measures the performance of these models and evaluates the relevant variables in heart attack prediction. One of the key findings of the study is the importance of factors like Maximum Heart Rate Achieved and Chest Pain Type in predicting heart attacks. The Support Vector Machine (SVM) model demonstrates the highest accuracy, precision, recall, and f1-score values, which are 85.53%, 82%, 95.35%, and 88.17% respectively. On the other hand, logistic regression is recommended as the bestperforming model overall based on its Mean Cross-Validation Score of 0.8239. This research emphasizes the potential of machine learning models such as SVM and logistic regression in enhancing heart attack prediction and offers valuable insights for the development of diagnostic techniques for heart disease (Audrey & Widianto, 2023).

	Accuracy	Precision	Recall	F1-score
Logistic Regression	82,89%	81,25%	90,70%	85,71%
KNN	81,58%	79,59%	90,70%	84,78%
SVM	85,53%	82,00%	95,35%	88,17%
Naïve Bayes	82,89%	81,25%	90,70%	85,71%
Decision Tree	80,26%	80,43%	86,05%	83,14%
Random Forest	82,89%	81,25%	90,70%	85,71%
XGBoost	77,63%	80,95%	79,07%	80,00%

Figure 5 Results of Various ML Classifiers

2.2.3. Heart Attack Prediction using Machine Learning Techniques

Author: J. Sharon Rose; P. Malin Bruntha; Salomi Salvias; Rajath M V; Bill Christ Mary M; Minni Jenifer D

Publisher: IEEE

Link: https://ieeexplore.ieee.org/document/10113045

Overview:

This study investigates the efficacy of various machine learning models, such as Support Vector Machine (SVM), Random Forest, Decision Tree, and Logistic Regression, in predicting heart attacks. The dataset used in the study was obtained from Kaggle and consists of fourteen heart-related attributes. Logistic regression yielded the highest accuracy of 82.4% and precision of 84.3%. The study demonstrates the potential impact of machine learning on prognosis and emphasizes the importance of early detection of heart issues. The suggested methods could be beneficial in the fields of fitness and medicine (Rose, et al., 2023).

Classifier	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F1-Score (%)
Logistic Regression	81.9	81.8	82.1	84.3	83
SVM	62.2	75.7	46.4	62.5	68.4
Decision Tree	81.4	84.8	78.5	82.3	83.5
Random Forest	77	75.7	78.5	80.6	78.1

Figure 6 Results of Various ML Classifiers

3. Proposed Solution

3.1. Explanation of the problem solution

One of the leading causes of mortality is cardiovascular disease. excessive blood pressure, excessive cholesterol, and a fast pulse rate are the basic causes of coronary disease. Tobacco uses and alcohol use are two separate risk factors for heart disease (Thummala & Baskar, 2022). Machine learning has numerous algorithms, and selecting the most effective machine learning algorithm is critical for accurate predictions and diagnosis.

The decision to use decision trees as a machine learning technique in this project is related to the nature of the dataset and its specific quality to address the problem at hand. Decision trees are particularly useful for this cardiovascular disease prediction model for the for different reasons such as interpretability, feature importance, handling nonlinearity , mixed data types, ease of explanations, identifying high risk subgroups, etc. This approach is consistent with the larger objective of enhancing cardiovascular health prevention techniques and patient outcomes.

3.2. Explanation of the AI algorithm/algorithms to be used.

A decision tree is a non-parametric supervised learning technique that may be used to perform classification and regression problems. It is a hierarchical tree with a root node, branches, internal nodes, and leaf nodes. Decision trees are used for classification and regression applications, resulting in simple models (Saini, 2023).It is a tool that has applications spanning several different areas. Decision trees can be used for classification as well as regression problems. The name itself suggests that it uses a flowchart like a tree structure to show the predictions that result from a series of featurebased splits. It starts with a root node and ends with a decision made by leaves (Saini, 2023).

3.2.1. Decision Tree Terminologies:

- **Root Node:** The first node of a decision tree, when the entire population or dataset begins to divide depending on various attributes or criteria (Saini, 2023).

- **Decision Nodes:** Decision nodes are nodes formed by the splitting of root nodes. These nodes in the tree indicate intermediate decisions or situations (Saini, 2023).

- **Sub-Tree:** A sub-section of a decision tree is known as a sub-tree, much as a subsection of a graph is known as a sub-graph. It represents a specific branch of the decision tree (Saini, 2023).

- **Pruning:** The removal or reduction of certain nodes in a decision tree in order to prevent overfitting and simplify the model (Saini, 2023).

- **Branch / Sub-Tree:** A branch or sub-tree is a part of the complete decision tree. Within the tree, it depicts a distinct route of decisions and results (Saini, 2023).

- **Parent and Child Node**: A node that is split into sub-nodes is known as a parent node in a decision tree, and the sub-nodes that emerge from it are known as child nodes. The parent node represents a decision or situation, and the child nodes indicate possible outcomes or further decisions based on that decision or condition (Saini, 2023).

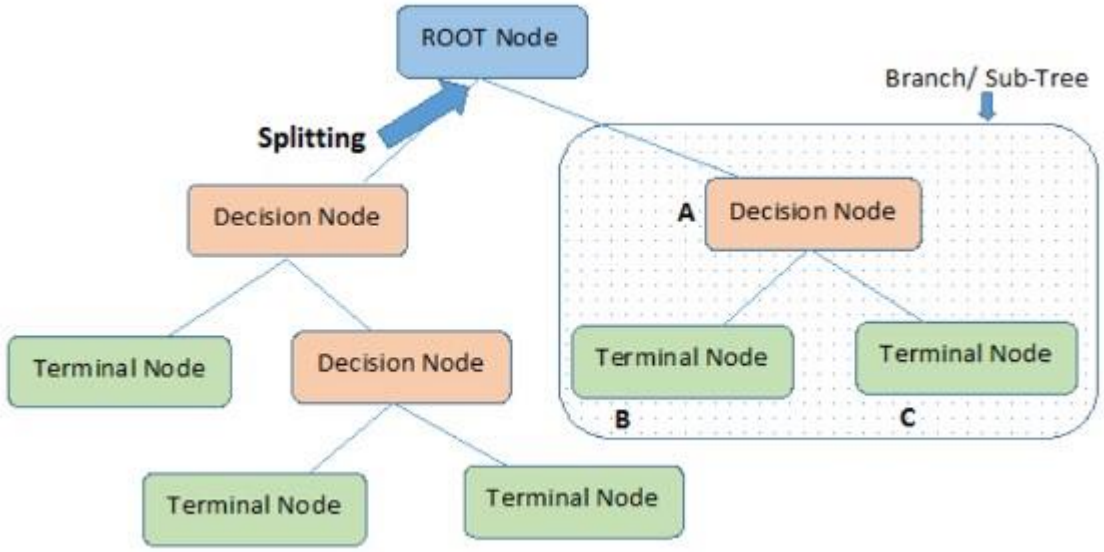

Figure 7 Decision tree Nodes

3.2.2. Attribute selection measures:

Attribute Selection Measures help in the selection of critical characteristics, preventing decision nodes from making arbitrary selections. They improve model accuracy by assessing the value of features, decreasing overfitting, and enhancing interpretability (Punia, 2022). The following are some of the most important attribute selection metrics used in decision trees:

- **Information Gain:** The amount of uncertainty reduced when dividing a dataset based on an attribute. It specifies how much information a feature provides about the class. Decision tree algorithms prioritize traits with the largest information gain (Javapoint, n.d.). It can be calculated using the formula given below:

Information Gain= Entropy(S)- [(Weighted Avg) *Entropy(each feature)

Figure 8 Formula for Information Gain

- **Entropy:** Entropy is a statistic used to quantify impurity in a particular characteristic. It denotes the unpredictability of data (Javapoint, n.d.). Entropy may be calculated as follows:

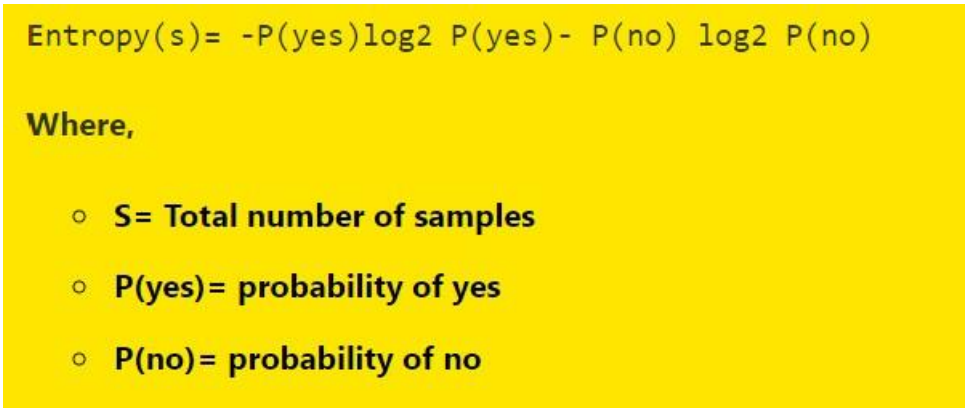

Figure 9 Formula for Entropy

3.3. Dataset

The dataset here used is derived from Kaggle. This dataset includes a wide range of attributes related to cardiovascular health and lifestyle decisions, such as age, gender, cholesterol levels, blood pressure, heart rate, physical activity duration, stress levels, etc (Mehfooz, 2021). All the attributes and their data types with are written in table below:

	Attributes	Description	Data type
1	Age	Age of the patient	Numeric
2	Sex	Gender of the patient	Numeric
3	cp	Chest pain type, 0 = Typical Angina, 1 = Atypical Angina, 2 = Non-anginal Pain, 3 = Asymptomatic	Numeric
4	trtbps	Resting blood pressure(in mm Hg)	Numeric
5	chol	Cholestoral in mg/dl fetched via BMI sensor	Numeric

6	fbs	(fasting blood sugar > 120 mg/dl) (1 = yes: 0 = No)	Numeric
7	restecg	Resting electrocardiographic results, 0 = Normal, 1 = ST-T wave normality, 2 = Left ventricular hypertrophy	Numeric
8	thalachh	Maximum heart rate achieved.	Numeric
9	Oldpeak	Previous peak	Float
10	Slp	Slope	Numeric
11	caa	Number of Major vessels	Numeric
12	thall	Thalium stress test result ~ (0,3)	Numeric
13	exng	Exercise induced angina (1 = yes: 0 = No)	Numeric
14	Output	Target variable	Numeric

Table 1 Dataset Description with data type (Mehfooz, 2021).

	age	sex	cp	trtbps	chol	fbs	restecg	thalachh	exng	oldpeak	slp	caa	thall	output
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0

Table 2 Dataset data frame

3.4. Pseudocode

START

IMPORT Libraries

LOAD Dataset

READ Dataset

SPLIT Dataset (Train and Test (7:3))

INITIALIZE Decision Tree Classifier

TRAIN the Decision Tree model

MAKE predictions on test data

EVALUATE model accuracy

USE the trained Decision Tree model for new predictions

END

3.5. Flowchart

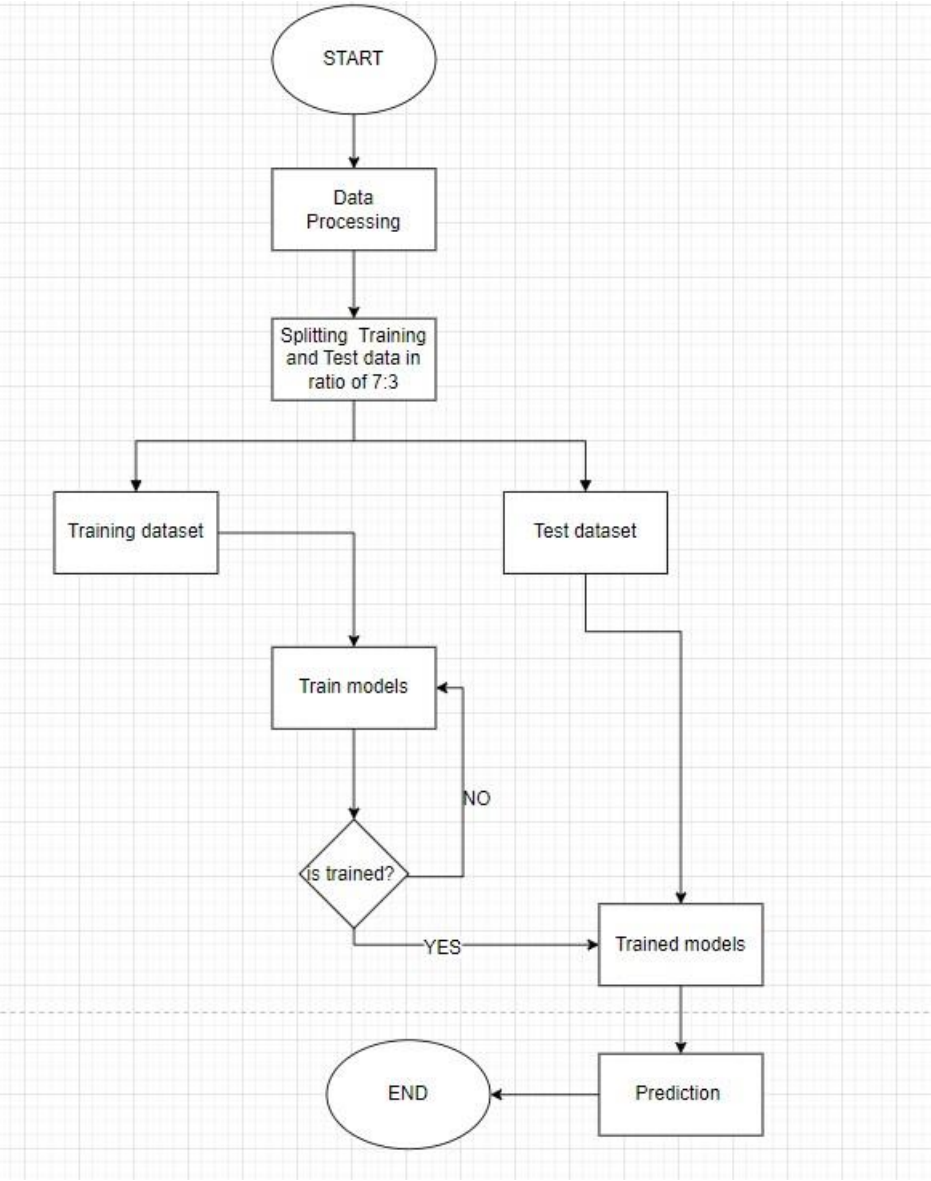

Figure 10 Flowchart

4. Conclusion

4.1. Analysis of the work done

Task	Status
Artificial Intelligence Research	Completed
Machine Learning Research	Completed
Research on AI Algorithms	Completed
Research on topic	Completed
Existing system review and analysis	Completed
Research on solution of problem solution	Completed
Flowchart and Pseudocode	Completed
Implementation	Not complete
Training data	Not complete
Testing project	Not complete

Table 3 Table of work done analysis.

4.2. How the solution addresses real world problems

The suggested method directly tackles real-world issues associated with cardiovascular diseases (CVDs) by utilizing a Decision Tree algorithm for forecasting heart attacks using machine learning. As, Cardiovascular diseases has been the major cause of mortality worldwide, claiming the lives of millions of people worldwide and for successful

intervention and prevention, early detection is crucial. The model wants to make a good guess about the things by looking at important parameters such as age, gender, cholesterol levels, blood pressure, and lifestyle factory by utilizing machine learning, specifically the interpretability and feature importance capabilities of Decision Trees.

This technology helps overcome the limitations of traditional diagnostic procedures and manual risk assessment. A decision tree approach can identify high-risk groups, allowing healthcare professionals to provide more individualized treatment to patients. Interpretability of decision trees allows a better understanding of the elements that control the prognoses, which allows for more informed clinical decisions. The goal is to improve preventive strategies for cardiovascular health, reduce premature deaths and improve public health outcomes.

4.3. Further work

The purpose of the project is to make sure that the development down the line will be completed, keeping in mind all the necessary tasks as listed in the table above, the status of all remaining works falls under ‘completed’ and ‘incomplete’ portion and the sole intend of making sure that the development of ‘Heart Attack Prediction System’ goes smoothly without any hitches.

To make sure that the development of the project is successful, various processes throughout the project are considered such as compiling the aggregate result of a successful review and a successful running of a system. The system in all makes use of proper ways of completion.

5. References

Audrey, J. & Widianto, M. H., 2023. *Heart Attack Prediction Using Machine Learning Classification Models.* Yogyakarta, Indonesia, IEEE.

Avcontentteam, 2023. *Machine Learning & AI for Healthcare in 2023.* [Online] Available at: https://www.analyticsvidhya.com/blog/2023/03/machine-learning-aifor-healthcare/
[Accessed 17 Dec 2023].

Frankenfield, J., 2023. *Artificial Intelligence (AI): What It Is and How It Is Used.* [Online] Available at: https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp
[Accessed 14 Dec 2023].

G, G., 2022. *Analysis of The Diagnostic Parameters of Heart Diseases and Prediction of Heart Attacks.* Bangalore, India, IEEE.

Javapoint, n.d. *Decision Tree Classification Algorithm.* [Online] Available at: https://www.javatpoint.com/machine-learning-decision-treeclassification-algorithm
[Accessed 19 Dec 2023].

Kanade, V., 2022. *What Is Artificial Intelligence (AI)? Definition, Types, Goals, Challenges, and Trends in 2022.* [Online] Available at: https://www.spiceworks.com/tech/artificial-intelligence/articles/whatis-ai/
[Accessed 15 Dec 2023].

Kanade, V., 2022. *What Is Machine Learning? Definition, Types, Applications, and Trends for 2022.* [Online]
Available at: https://www.spiceworks.com/tech/artificial-intelligence/articles/whatis-ml/
[Accessed 17 Dec 2023].

Mehfooz, F., 2021. *HeartAttack prediction with 91.8 % Accuracy.* [Online]
Available at: https://www.kaggle.com/code/fahadmehfoooz/heartattack-predictionwith-91-8-accuracy [Accessed 18 Dec 2023].

NCD Alliance, 2023. *Cardiovascular Diseases (CVD).* [Online] Available at: https://ncdalliance.org/news-events/news
[Accessed 17 Dec 2023].

Punia, J., 2022. *Decision Tree Algorithm.* [Online]

Available at: https://www.scaler.com/topics/decision-tree-algorithm/ [Accessed 19 Dec 2023].

Rose, J. S. et al., 2023. *Heart Attack Prediction using Machine Learning Techniques.* Coimbatore, India, IEEE.

Saini, A., 2023. *Decision Tree Algorithm – A Complete Guide.* [Online] Available at: https://www.analyticsvidhya.com/blog/2021/08/decision-treealgorithm/#What_is_a_Decision_Tree? [Accessed 18 Dec 2023].

Sharma, R., 2022. *What are Machine Learning Applications? Top 10 Industry and Real-World Use Cases.* [Online] Available at: https://emeritus.org/blog/machine-learning-what-are-machinelearning-applications/ [Accessed 17 Dec 2023].

Thummala, G. S. R. & Baskar, R., 2022. *Prediction of Heart Disease using Decision Tree in Comparison with KNN to Improve Accuracy.* Chennai, India, IEEE.

University of Oxford, 2023. *AI tool could help thousands avoid fatal heart attacks.* [Online] Available at: https://www.ox.ac.uk/news/2023-11-13-ai-tool-could-helpthousands-avoid-fatal-heart-attacks [Accessed 19 Dec 2023].

World Health Organization, 2023. *Cardiovascular diseases.* [Online] Available at: https://www.who.int/health-topics/cardiovasculardiseases#tab=tab_1 [Accessed 17 Dec 2023].