Utilizing Artificial Intelligence Techniques for Predicting Cases of Coronary Disease: An Investigative Approach

Lakka Jahnavi

Department of Computer Science Sri Venkateswara University, Tirupati

***Abstract*—** *Coronary disease stands as one of the most significant human health challenges worldwide, profoundly impacting human lives. Cardiovascular disorders, including heart-related ailments, have been responsible for a considerable number of deaths globally over the past few decades, emerging as the deadliest disease not only in India but also worldwide. Accurate and timely diagnosis of coronary disease is crucial for preventing cardiovascular failure and ensuring effective treatment. Thus, there is a pressing need for robust, precise, and efficient systems to diagnose such infections promptly for appropriate treatment. In this study, we utilized the Heart Stalog dataset obtained from the UCI repository, employing Neural Networks and Logistic Regression algorithms to accurately predict the occurrence of coronary disease. The proposed decision support system based on Neural Networks and Logistic Regression will aid healthcare professionals in efficiently identifying heart patients. Logistic Regression emerged as the superior model among the two algorithms, achieving an overall accuracy rate of 91.54%. Our findings demonstrate the superior performance of Logistic Regression over Neural Networks in terms of precision. Developing accurate and computationally efficient classifiers for clinical applications remains a significant challenge in Machine Learning.*

I. INTRODUCTION

In recent years, there has been a significant surge in interest in analyzing clinical data, as healthcare organizations recognize the potential of leveraging patient data from various clinical systems to improve healthcare delivery and management of clinical datasets. Extracting insights from such data requires advancements from the realms of Data Mining, Machine Learning, Artificial Intelligence, and Data Visualization.

Healthcare institutions are now generating vast amounts of data, posing challenges in data management. Hospitals have accumulated extensive patient data and medical records, prompting the need for data mining to uncover patterns and associations that can inform effective decision-making. Clinical data mining is pivotal for extracting valuable clinical insights from medical datasets.

This urgency stems from various critical health issues such as heart disease, liver failure, kidney complications, nerve damage, and vision impairment. Early identification of diabetes is among the significant clinical challenges. The heart, being the central organ in the human body, plays a vital role, and any impairment can affect other crucial body functions, necessitating cardiac health assessments.

Coronary artery disease (CAD) is regarded as one of the most complex and life-threatening human diseases globally. It impedes the heart's ability to pump an adequate amount of blood to fulfill the body's normal functions, leading to heart failure. According to the World Health Organization (WHO), an estimated 17 million people die annually from cardiovascular diseases, including coronary failures and strokes.

Symptoms of heart disease include shortness of breath, fatigue, swollen feet, and other associated signs indicative of cardiovascular or noncardiac abnormalities. Early-stage diagnostic methods for identifying heart disease were complex, and their subsequent complications have significantly impacted quality of life. Diagnosis and treatment of heart disease remain challenging, particularly in non-industrialized countries, due to limited access to medical resources and a shortage of healthcare professionals, affecting accurate diagnosis and treatment of patients.

Accurate and timely diagnosis of coronary artery disease risk in patients is crucial for reducing their associated risks of severe heart issues and improving heart health outcomes. [1]. [7]. [6].

II. CLASSIFICATION SYSTEM

Course of action is the way toward finding a model or a limit that depicts and perceives data classes and thoughts, to use the model to predict the classes of things whose class mark isn't known. Data request should be visible as a two-stage measure:

learning step in which a classifier is built depicting a predestined course of action of classes or thoughts by separating the readiness set included informational index tuples and their connected names. In the resulting advance model is used for request by first surveying the perceptive accuracy of classifier worked during the underlying advance. It is done using the test data. The precision of classifier on a given test set tuples is level of tuples that are precisely requested by the classifier. If the accuracy is over some sufficient level, the classifier can be used to expect future tuples whose class mark isn't known.

Portrayal is a kind of data assessment that can be used to create models portraying huge data classes. Game plan is a data mining system used to predict bundle investment for data models. It is one of the critical systems in data mining and is used in various applications, for instance, plan affirmation, disease assurance, client relationship the chiefs, and assigned displaying. The goal of the portrayal estimations is to fabricate a model from a lot of planning data whose target class names are known and subsequently this model is used to bunch disguised cases [2][3].

Course of action is the most normal and most renowned data mining procedures. Game plan maps data into predefined social occasions or classes. It is ordinary insinuated as managed learning considering the way that the classes are settled preceding checking the data out. Game plan is the way toward finding a model that perceives data classes, to use the model to anticipate the class of things whose class name is dark. The decided model relies upon the assessment of a lot of getting ready data. Informational indexes are rich with concealed information that can be used for watchful dynamic.

Building precise and useful classifiers for immense data bases is one of the crucial tasks of data mining and AI research. Building effective request structures is one of the central tasks of data mining.

A wide extent of kinds of assortment systems have been proposed recorded as a printed copy that combine Decision Trees, Naive-Bayesian procedures, Neural Networks, Logistic Regression, Support Vector Machines (SVM) and K-Nearest Neighbor, and so forth.

III. METHODOLOGY

At the present time, made sense of about directed learning procedures like Random Forest and Logistic Regression system models for Heart Stalog sickness characterization issue.

3.1 Artificial Neural Networks

An Artificial Neural Network (ANN) is a computational model awakened in the working of the human frontal cortex do portrayal. ANNs have been comprehensively used in model affirmation, talk affirmation and remedial finding, fault acknowledgment, issue assurance, robot control, and PC vision. ANNs have the limit of appropriated information, equal dealing with, variation to non-basic disappointment and self-affiliation. It is made by a set out of phony neurons (known as planning units) that are interconnected with various neurons, essential taking care of parts called counterfeit neurons [4] [6]. In taking care of information, the planning parts in an ANN work all the while and overall thusly to natural neurons. Each affiliation has a weight related that addresses the effect from one neuron on the other. The loads connected with the data layer, concealed and yield layer affiliations are instated with minimal sporadic regard which is under 0.1. Along these lines, the neurons of the yield layer make a result from the hid layer and the information layer according to a sigmoidal commencement work showed up as Equation: 1/1+e-t (1)

A brain framework has three layers in its construction. First layer is input layer which is clearly partner with external universes; second layer is of covered unit where computation is done by limit gave, the last layer is yield layer from where we get yield. Learning in brain frameworks is taken care of as synaptic loads between neurons. The framework spreads the data from one layer to another until the yield data is delivered. If the frameworks is multi-facet perceptron with Backpropogation estimation and the yield isn't exactly equivalent to the hankering yield, by then a not entirely settled and multiplied backward through the framework.

ANN is an arrangement of associated neurons composed in layers:

- **Input layer:** Input layer is an element vector of issue, with various neurons equivalent to the quantity of factors of the issue, it carries the underlying information into the framework for further handling by ensuing layers of fake neurons.
- **Hidden layer:** A Hidden layer gets data from information layer and plays out every one of the procedures it, where fake neurons take in an arrangement of weighted sources of info and create a yield through an actuation work.

Concealed layers may shift from system to organize that we chose. Number of concealed layers relies on the nature and size of issue.

- **Output layer:** Output layer gets prepared data from shrouded layers and gives the yield to the outer client.

3.2 Logistic Regression (LR)

LR is considered as the standard genuine method for managing showing twofold data [4]. The central mathematical thought that underlies determined backslide is the logit — the typical logarithm of a possibilities extent. It's everything except an unrivaled choice for an immediate backslide which assigns a straight model to all of the class and predicts unnoticeable events basing on bigger part vote of the models. Overall, key backslide is proper for portraying and testing hypotheses about associations between an outright outcome variable and something like one full scale or diligent marker factors. During assumption, instead of predicting the point measure of the genuine event, it's everything except a model to expect the possibilities of its occasion. In two class issue for example, whenever the odds are more unmistakable than half, then the case is given out to the class doled out as "1" for YES and "0" for "YES" and "NO" in light of everything.

IV. EXPERIMENTAL RESULTS

The trial was executed the two calculations (Logistic Regression and Random Forest) utilizing WEKA. WEKA represents Waikato Environment for Knowledge Analysis. WEKA is made by analysts at the University of Waikato in New Zealand. The product is written in the Java language and contains a GUI for collaborating with information documents. WEKA additionally gives the graphical UI of the client and gives numerous offices. WEKA is a cutting-edge office for creating AI (ML) methods and their application to true information mining issues. WEKA executes calculations for information pre-preparing, grouping, relapse and bunching and affiliation rules. It likewise incorporates perception devices. We have considered the Heart statlog Disease information from UCI Machine Learning Repository datasets [8], for evaluating the efficiency and sufficiency of Logistic Regression and Random Forest frameworks. The dataset comprises of 270 records and 14 ascribes of exchanges and have two classes to be specific Absent (150) and Present (120) The characteristic data information is dense in figure-1. The standard dataset is apportioned into two sets (70% and 30%), one for planning and another set for testing.

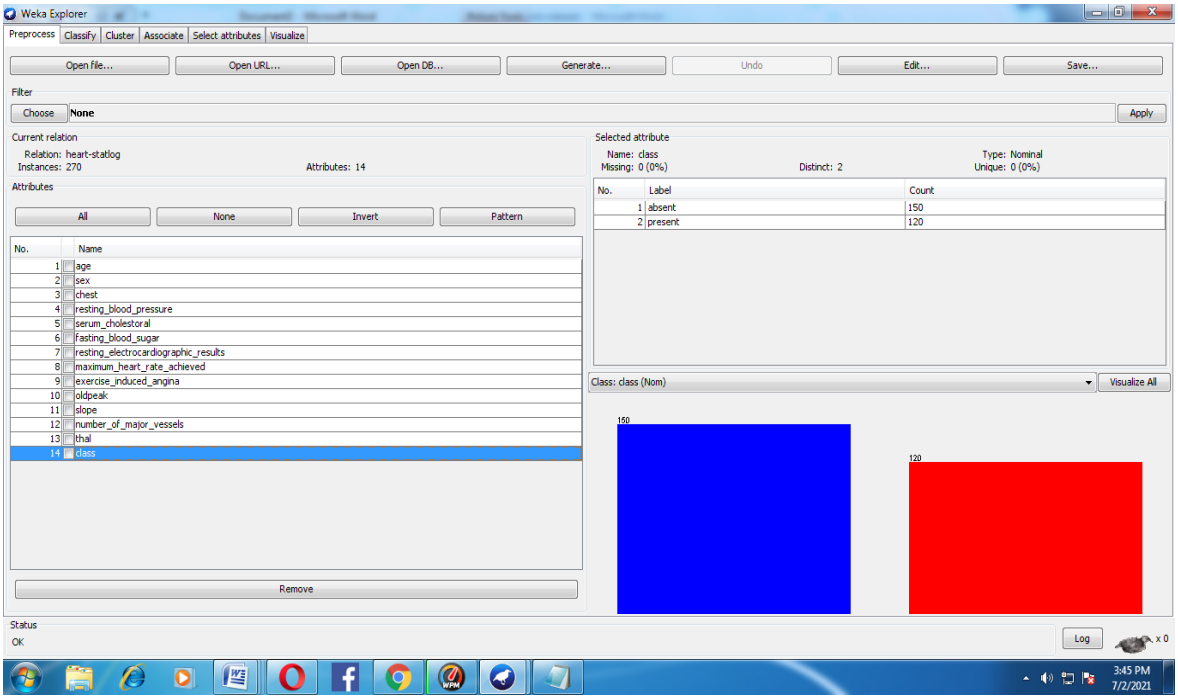

Figure-1: Summary of the Heart Stalog dataset

We have applied the analysis on the test information after pre preparing utilizing two forecast models. We assess our two models utilizing diverse execution measurements like exactness, accuracy and Recall, the Experimental outcomes are appeared in the table-1 and same appeared in the Figure-2.

Table-1
Performance of classifiers

Algorithm	Accuracy	precision	Recall
Neural Networks	88	88	87.78
Logistic Regression	91.54	91	91.23

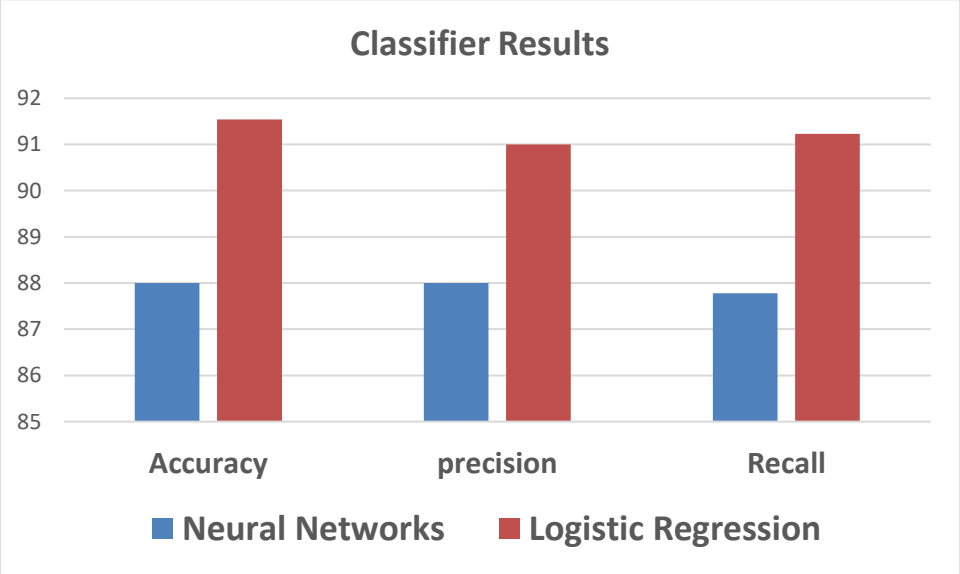

Figure-2: Performance of Classifier

We see in the Figure-2, the presentation of the Logistic Regression calculation has achieved 91.54% exactness and Neural Network has accomplished 88%. As the outcome from examination among the two calculations, we locate that most noteworthy exactness of Classification model is Logistic Regression (91.54%). Exactly when diverged from accuracy and review are moreover higher in the Logistic Regression model when contrasted with Neural Networks models.

V. CONCLUSION

The abundance of clinical datasets available for various data mining and AI techniques underscores the importance of enhancing the accuracy and efficiency of disease diagnosis. The objective of this research is to demonstrate how classifying Heart Stalog disease categories from publicly available raw clinical datasets can assist physicians in reaching precise diagnoses to predict the presence or absence of heart disease. Upon evaluating the results, Logistic Regression emerged with the highest prediction accuracy of 91.54%. This model proves to be the most effective in predicting patients with coronary disease. Thus, the proposed Logistic Regression Classifier approach offers a reliable method for both prediction and diagnosis.

REFERENCES

[1] HeonGyu Lee, Ki Yong Noh, KeunHoRyu, "Mining Biosignal Data: Coronary Artery Disease Diagnosis using Linear and Nonlinear Features of HRV," LNAI 4819: Emerging Technologies in Knowledge Discovery and Data Mining, pp. 56-66, May 2007.

[2] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann, 2005.

[3] J.Han and M.Kamber,"Data Mining concepts and Techniques", the Morgan Kaufmann series in Data Management Systems, 2nd ed.San Mateo, CA; Morgan Kaufmann, 2006.

[4] N.Michael, "Artificial Intelligence – A Guide to Intelligent Systems", 2nd Edition, Addison Wesley 2005

[5] P.N.Tan, M.Steinbach and V.Kumar "Introduction to Data Mining", A: Addision-Wesley, 2005.

[6] Sitar-Taut, V.A., et al., Using machine learning algorithms in cardiovascular disease risk evaluation. Journal of Applied Computer Science & Mathematics, 2009.

[7] "The Atlas of Heart Disease and Stroke", http://www.who.int/cardiovascular_diseases/res ources/atlas/en/

[8] UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/.