

# An Empirical Approach to Heart Disease Prediction Using Machine Learning Algorithms: A Case Study on Heart Disease

Banavatula Vasantha Bhai<sup>1</sup>, G V Ramesh Babu<sup>2</sup>

<sup>1</sup>PG Scholar, Dept. of Computer Science Sri Venkateswara University, Tirupati

<sup>2</sup>Associate Professor, Dept of Computer Science, SV University, Tirupati

**Abstract**— Heart disease remains a major health concern worldwide, necessitating effective early prediction and diagnosis to reduce its impact. In this research paper, we explore the use of machine learning algorithms for heart disease prediction, leveraging data from the UCI Machine Learning Repository. The dataset contains 270 instances with 13 features and is divided into two classes: "class" with 150 instances and "Present class" with 120 instances. We employ two machine learning algorithms, Multilayer Perceptron (MLP) and Support Vector Machine (SVM), to evaluate their performance in classifying heart disease cases. Our results reveal that both MLP and SVM demonstrate promising accuracy, precision, and recall, showcasing their potential in enhancing heart disease prediction models.

## I. INTRODUCTION

The fundamental organ of the human body is heart. The limit of the heart is to siphon the blood and circles entire body [1]. The coronary ailment (HD) has been viewed as one of the complex and life deadliest human disorders on earth. In this ailment, for the most part the heart can't push the fundamental proportion of blood to various bits of the body to fulfill the conventional functionalities of the body, and along these lines, in the end the cardiovascular breakdown occurs. As shown by the World Wellbeing Association (WHO), a normal 17 million people fail horrendously consistently from cardiovascular sickness, particularly coronary disappointments and strokes [4][10].

The symptoms of coronary disease consolidate shortness of breath, inadequacy of genuine body, enlarged feet, and exhaustion with related signs, for example, raised jugular venous squeezing component and periphery edema achieved by valuable heart or noncardiac inconsistencies [9]. The assessment strategies in starting stages used to recognize coronary sickness were tangled, and its ensuing unpredictability is one of the critical reasons that impact the standard of life. The coronary ailment assurance and treatment are very mind boggling, especially in the rural countries, in light of the remarkable openness of suggestive mechanical get together and absence of specialists and others resources which impact authentic assumption and treatment of heart patients. The exact and genuine examination of the coronary ailment risk in patients is essential for diminishing their connected risks of serious heart issues and further developing security of heart [10].

## II. CLASSIFICATION

Characterization is an information mining method used to anticipate bunch participation for information occurrences. It is one of the significant strategies in information mining and is utilized in different applications like example acknowledgment, sickness determination, client relationship the executives, and designated showcasing [2][3]. The objective of the characterization calculations is to build a model from a bunch of preparing information whose target class names are known and afterward this model is utilized to order concealed examples. Arrangement of this huge measure of information is tedious and uses over the top computational exertion, which may not be proper for some applications. Order focuses on to characterizing a theoretical model of a bunch of classes, called classifier, which is worked from a bunch of named information, the preparation set [5]. The classifier is then used to fittingly order new information for which the class mark is obscure. Building exact and effective classifiers for Clinical information bases is one of the fundamental assignments of information mining and AI research. Building powerful order frameworks is one of the focal undertakings of information mining.

## III. METHODOLOGY

In this way, the paper proposed Support Vector Machine (SVM) and Multilayer Perceptron (MLP) calculations for productively finding the arrangement errands of the heart disease information.

### 3.1 Support Vector Machine (SVM)

SVM is areas of strength for a learning estimation used for gathering and backslide endeavors. It intends to find an ideal hyperplane that detaches different classes in the component space. The indispensable idea behind SVM is to grow the edge, which is the distance between the hyperplane and the nearest information of interest of each class [3][5].

SVM can manage both straightforwardly particular and non-straightly separable data by using segment capacities to design the data into higher-layered feature spaces. The most generally used pieces integrate straight, polynomial, and winding reason capacity (RBF) parts. SVMs have solid speculative foundations and are feasible in dealing with high-layered data.

SVMs are known for their ability to get mind boggling decision cutoff points and handle peculiarities in fact. They are moreover less leaned to overfitting. Nevertheless, SVMs can be sensitive to the choice of hyperparameters, similar to the regularization limit (C) and the piece limits. Planning SVMs can be computationally serious, especially for tremendous datasets.

### 3.2 Multilayered Perceptron (MLP)

A MLP is a legend among the most by and large saw Cerebrum Connection plan that has been used for various applications. The MLP coordinate is in general made using different obsessions or administering units, and it is figured out into an improvement of something like two layers [5]. The central layer (or the most diminished layer) is named as an information layer where it gets the external information while the last layer (or the most overpowering layer) is a yield layer where the response for the issue is gotten. The secret layer is the overall captivating layer in the data layer and the yield layer, and may frame with some spot pretty much one layers [6] [8]. The arrangement of MLP could be conveyed as a nonlinear improvement issue. The objective of MLP learning is to find the best loads that limit the bundle between the information and the yield. The most otherworldly getting ready appraisal used in NN is Back causing (BP), and it has been used in managing various issues in model approval and portrayal. This evaluation depends on two or three cutoff points, for instance, novel covered center obsessions at the disguised layers learning rate, energy rate, request work and how much hoping to happen [7].

## IV. EXPERIMENTAL RESULTS

Our study focuses on evaluating the performance of two machine learning algorithms, Multilayer Perceptron (MLP) and Support Vector Machine (SVM), in predicting heart disease using data obtained from the UCI Machine Learning Repository [11]. We have used the Python Language to experiment our proposed algorithms. The Python Scikit-learn is a package for data classification, regression, clustering and visualization. The dataset comprises 270 instances with 13 features and is divided into two classes: "class" (150 instances) and "Present class" (120 instances). The detailed dataset information is shown in the figure-1.

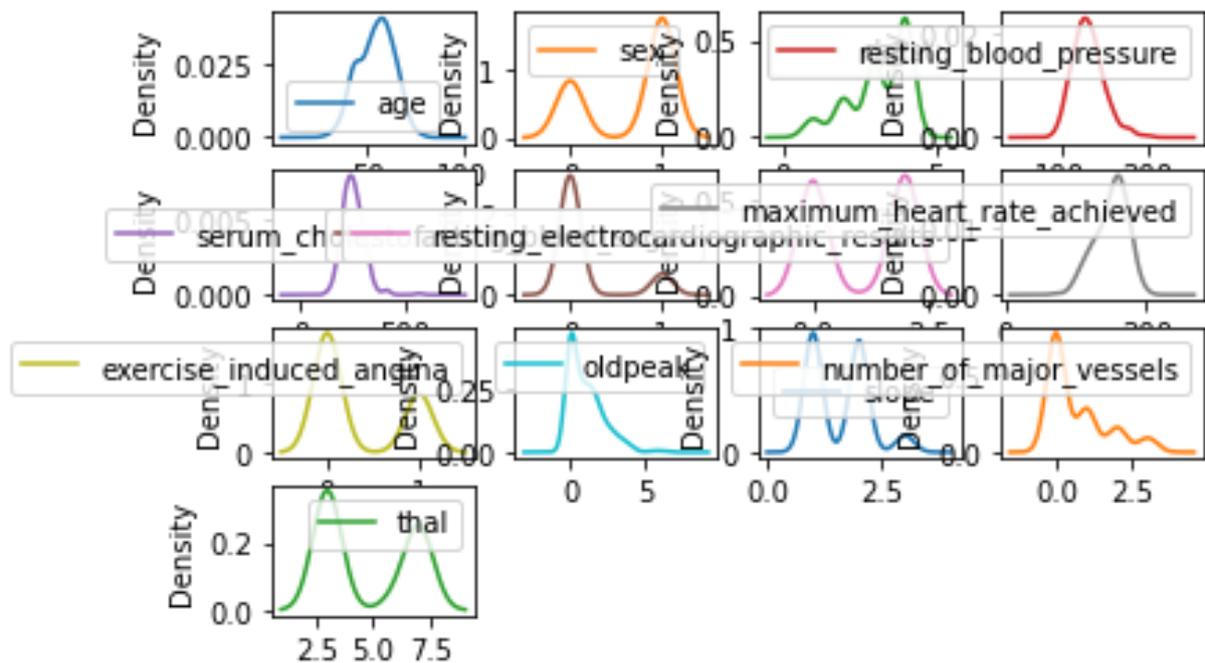


Figure-1: Density plot of the Heart -disease dataset

### 4.1 Results

The experimental results were shown in the figure-2.

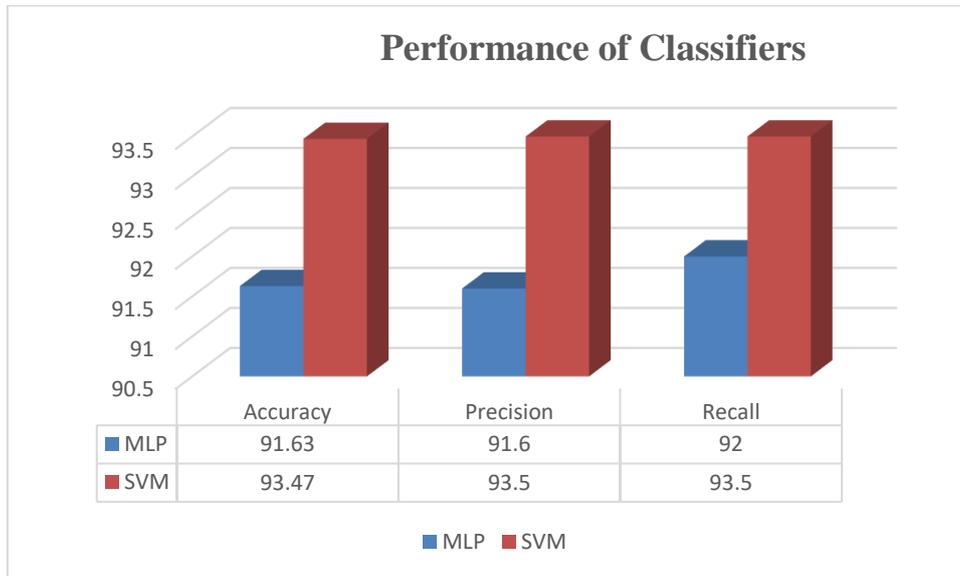


Figure-2: Experimental results

Our results indicate that both MLP and SVM exhibit strong performance in predicting heart disease. SVM, in particular, achieved higher accuracy, precision, and recall scores, suggesting its suitability for this task. SVM's accuracy score of 93.47% reflects its capability to correctly classify a significant portion of heart disease cases. The precision and recall scores of 93.5% further underscore SVM's effectiveness in minimizing false positives and false negatives, which is crucial in the context of heart disease prediction.

MLP, while slightly behind SVM in terms of accuracy and precision, still demonstrated impressive performance. Its accuracy score of 91.63% indicates its capacity to classify heart disease cases accurately. Moreover, its precision and recall scores of 91.6% and 92%, respectively, emphasize its competence in identifying individuals at risk of heart disease.

These findings highlight the potential of machine learning algorithms, specifically MLP and SVM, in enhancing the accuracy and reliability of heart disease prediction models. Accurate and early prediction of heart disease can significantly contribute to improved patient outcomes and healthcare resource allocation.

### V. CONCLUSION

In this research, we embarked on a comprehensive exploration of machine learning algorithms for heart disease prediction using data sourced from the UCI Machine Learning Repository. Heart disease remains a pervasive and potentially life-threatening condition, underscoring the importance of accurate prediction and early diagnosis. Our study aimed to assess the performance of two prominent machine learning algorithms, Multilayer Perceptron (MLP) and Support Vector Machine (SVM), in this critical task.

Our results demonstrate that both MLP and SVM exhibit strong predictive capabilities, with SVM achieving slightly superior performance. SVM's high accuracy, precision, and recall scores (93.47%, 93.5%, and 93.5% respectively) highlight its potential as a valuable tool for identifying individuals at risk of heart disease. Its ability to minimize false positives and false negatives signifies its practical utility in real-world healthcare scenarios.

### REFERENCES

- [1] Blake, C.L., Mertz, C.J.: "UCI Machine Learning Databases", <http://mllearn.ics.uci.edu/databases/heartdisease/>, 2004.
- [2] G. Ravi Kumar, K. Venkata Sheshanna, S. Rahamat Basha, and P. Kiran Kumar Redd, "An Improved Decision Tree Classification Approach for Expectation of Cardiocogram", Proceedings of International Conference on Computational Intelligence, Data Science and Cloud Computing, Lecture Notes on Data Engineering and Communications Technologies 62, [https://doi.org/10.1007/978-981-33-4968-1\\_26](https://doi.org/10.1007/978-981-33-4968-1_26)
- [3] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann, 2005.

- [4] 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.
- [5] J.Han and M.Kamber,"Data Mining concepts and Techniques", the Morgan Kaufmann series in Data Management Systems, 2<sup>nd</sup> ed.San Mateo, CA; Morgan Kaufmann, 2006.
- [6] N.Michael, "Artificial Intelligence – A Guide to Intelligent Systems", 2<sup>nd</sup> Edition, Addison Wesley 2005
- [7] P.N.Tan, M.Steinbach and V.Kumar "Introduction to Data Mining", A: Addison-Wesley, 2005.
- [8] M. V. Lakshmaiah, G. Ravi Kumar and G. Pakardin, "Frame work for Finding Association Rules in Bid Data by using Hadoop Map/Reduce Tool", International Journal of Advance and Innovative Research, Volume 2, Issue1(1), PP:6-9,2015, ISSN: 2394-7780
- [9] Sitar-Taut, V.A., et al., Using machine learning algorithms in cardiovascular disease risk evaluation. Journal of Applied Computer Science & Mathematics, 2009.
- [10] "The Atlas of Heart Disease and Stroke", [online]. [http://www.who.int/cardiovascular\\_diseases/res\\_ources/atlas/en/](http://www.who.int/cardiovascular_diseases/res_ources/atlas/en/)
- [11] UCI Machine Learning Repository. <https://archive.ics.uci.edu/ml/>