# Comparative Survey of Machine Learning Algorithms for Heart Disease Prediction: Trends, Challenges, and Future Directions.

Rivindi Kalasing Ekanayake. (MSc in Advanced Computer Science-24012043)

**Abstract**

## 1. Introduction

Heart disease remains a leading cause of death worldwide, claiming millions of lives an estimated around 17 million deaths each year by claiming 32%, and getting considerable strain on healthcare systems globally(*Cardiovascular diseases (CVDs)* n.d.). The ability to identify and predict cardiovascular risk early offers numerous benefits, such as lives can be saved, healthcare resources allocated more efficiently and effectively, and patients can have longer, healthier lives with their loved once. Yet our current diagnostic toolkit, despite its polish, it has real limitations. All the clinical examinations rely on human judgment, ECGs can miss early warning signs of the cardiovascular, and even advanced imaging sometimes fails to reveal subtle pathology before it becomes critical stage. Complex interactions between risk factors often go undetected simply because the human eye cannot process that level of data layers.

Machine learning offers a different approach to the several real-world practices, in several industries. In healthcare also, rather than demanding clinicians hold dozens of variables in mind simultaneously, ML algorithms can absorb patterns from large patient datasets automatically and prediction due to ability to analyses large datasets, identify non-linear relationships, and extract predictive patterns of the data beyond human capability. Classical statistical approaches are outperformed by ML models in risk prediction, symptom classification, and ECG analysis(Krittanawong *et al.* 2020)

### 1.1. Significance of the comparative survey

These ML based algorithms can automatically learn from clinical features for the patients, such as blood pressure, cholesterol levels, chest pain type, glucose level, and ECG parameters and can help

clinicians make earlier, more reliable decisions. Due to the high availability of datasets such as the UCI Cleveland Heart Disease Dataset(Detrano *et al.* 1989), the Statlog Heart Dataset, and larger population datasets like the Framingham Heart Study (D'Agostino, Vasan, Pencina, Wolf, Cobain, Massaro & Kannel 2008) has speeded up the adoption of ML models in cardiac prediction research. most of the studies demonstrate that algorithms such as Support Vector Machines (SVM), Random Forests (RF), Logistic Regression (LR), Artificial Neural Networks (ANN), and hybrid methods such as XGBoost achieve high predictive accuracy and robust performance (Khan, Nawi, Shahzad, Ullah, Mushtaq, Mir & Aamir 2017).

However, despite rapid growth in ML-based cardiovascular analytics and prediction algorithm developments, research findings are scattered, vary across datasets, and often lack standardisation in preprocessing methods, evaluation metrics, and validation strategies. Many studies rely on small, outdated datasets and do not address interpretability concerns of the data set, which limits clinical applicability. Integrating existing evidence through a comprehensive review is essentiall to evaluate ML algorithmic performance in heart disease prediction, reveal current gaps in the field, and identify future research opportunities.

### 1.2. Aims and objectives

This study aims to provide a detailed comparative survey of machine learning algorithms used for heart disease prediction, focusing on emerging trends, challenges, and potential future research developments in the healthcare field.

**Research questions**

- How do various machine learning algorithms perform relative to one another across accuracy, interpretability, computational efficiency, and robustness in cardiac disease prediction?
- What knowledge gaps, methodological challenges, and limitations are evident in current literature, and what approaches could prospective studies adopt to overcome them?

**Objectives of the Study**

- To systematically survey and classify existing literature on machine learning methodologies employed in cardiac disease prediction.

- To contrast conventional ML techniques (e.g., LR, SVM, RF, KNN, NB) with approaches (e.g., ANN, ensemble models, deep learning) utilising established performance indicators.
- To critically evaluate methodological deficiencies, potential limitations, and interpretability challenges evident in current selected studies.
- To propose future research tpathways that enhance the efficacy, reliability, and clinical applicability of ML models in cardiovascular disease prediction.