COMM022 Coursework

Machine Learning Algorithms

Proposal of

Comparative Analysis of Machine Learning Algorithms for Heart Disease Prediction.

Rivindi K. K. Ekanayake

24012043

MSc in Advanced Computer Science

Dr. Neil Buckley

Liverpool Hope University

Heart disease continues to be a significant contributor to mortality rates worldwide, making it critical to identify cases promptly and begin treatment early. This study examines how different computational learning methods perform when analyzing heart disease risk, drawing from patient information available in the UCI Heart Disease dataset (András Jánosi, William Steinbrunn, Matthias Pfisterer & Robert, 1988). The research evaluates multiple classification-based models, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbour (KNN) to determine which approach delivers the strongest results when measured by accuracy, precision, recall, and F1-score metrics. The outcomes should offer actionable guidance for both healthcare professionals and computational specialists working to strengthen diagnostic prediction systems and refine how technology supports medical decision-making in clinical settings.

Introduction to the problem statement

Using statistics and automated systems to forecast heart disease has grown into a prominent research area, driven by increasing access to patient datasets in the public domain and improvements in machine intelligence technologies. Conventional diagnostic approaches typically depend on clinician judgment and consume substantial time. Machine Learning based systems present a feasible solution capable of recognising trends across extensive patient records while facilitating quicker, effective and less subjective conclusions (Obehi Winnifred Ikpea & Dr David Han., 2022). Earlier research has shown promise in applying ML to cardiovascular risk assessment, although systematic comparisons between different computational models with standardized data preparation and assessment criteria have been scarce. This research study fills this gap of the studies by implementing various computational techniques using ML algorithms on identical data set while maintaining consistent experimental parameters throughout the study (Munmun, Z.S., Akter, S. & Parvez, C.R., 2025).

Current individual computational models have produced encouraging outcomes for the prediction generations foe the heart diseases, this study determine which method proves most reliable for heart disease forecasting in practical cases in healthcare environments remains uncertain. The variety in data preparation methods, normalization techniques, and assessment approaches frequently produces conflicting findings. This research study focuses to deliver a thorough side-by-side performance assessment of established computational techniques with ML algorithms, leveraging a unified dataset and consistent assessment methodology.

Objectives

- To carry out a comparative analysis of machine learning algorithms for heart disease prediction using the UCI Heart Disease dataset.
- To preprocess and clean the UCI dataset for effective model training.
- To implement and train selected ML algorithms (Logistic Regression, Random Forest, SVM, and KNN).
- To evaluate each algorithm's performance using standard metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.
- To identify the most efficient and accurate algorithm for heart disease prediction.
- To provide recommendations for practical implementation in healthcare data analysis.

Methodology and Expected Outcomes

The research study will utilize a numerical research design grounded in structured experimentation. Clinical information will be obtained from the UCI Heart Disease Dataset, which encompasses patient measurements including age, gender, lipid concentrations, arterial pressure, and cardiac response capacity. Initial data preparation will involve resolving incomplete entries, standardizing values across features, screening for relevant variables, and partitioning into development (80%) and validation (20%) subsets. Four computational techniques will be constructed using Python frameworks (Scikit-learn, NumPy, Pandas, Matplotlib). System assessment will apply repeated sampling validation to strengthen the reliability of generated outcomes (Rimal, Y., Sharma, N., Paudel, S., Alsadoon, A., Koirala, M.P. & Gill, S., 2025). Quantitative evaluation indicators, encompassing classification accuracy, positive predictive value, sensitivity, harmonic mean of precision and recall, and curve analysis under the probability threshold will enable systematic comparison across models. The technique demonstrating superior equilibrium between predictive accuracy and ease of explanation will emerge as the preferred option for real-world deployment.

The complete implementation and study are projected to pinpoint the computational approach that generates the most robust and dependable forecasts for identifying heart conditions of the patients. Additionally, it will underscore how critical preparatory data processing and targeted variable selection are for strengthening system capabilities and reliability of the outcome. The results will advance the creation of more streamlined technological resources that facilitate clinical judgment

in medical practice. This study project employs a de-identified, publicly distributed information source containing no sensitive personal markers, thereby preventing exposure of private information relates to the ethics of data usage. All materials will be managed with care and dedicated exclusively to knowledge advancement and scholarly inquiry, in accordance with organizational research conduct standards to practice the proper ethical practice handling.

References

András Jánosi, William Steinbrunn, Matthias Pfisterer & Robert, 1988. *Heart Disease (UCI Machine Learning Repository).* [Online]
Available at: https://archive.ics.uci.edu/dataset/45/heart%2Bdisease
[Accessed 04 11 2025].

Jánosi, A., Steinbrunn, W., Pfisterer, M. & Detrano, R., 1988. Heart Disease [Dataset]. *UCI Machine Learning Repository, University of California, Irvine..*

Mohammad Belal Aziz, Syed Wajahat Abbas Rizvi, 2025. Comparative Analysis Of Machine Learning Algorithms For Heart Disease Prediction. *International Journal of Computer Applications,* Volume 187.

Munmun, Z.S., Akter, S. & Parvez, C.R., 2025. Machine Learning-Based Classification of Coronary Heart Disease: A Comparative Analysis of Logistic Regression, Random Forest, and Support Vector Machine Models. *Open Access Library Journal,* Volume 12, pp. 1-12.

Obehi Winnifred Ikpea & Dr David Han., 2022. Performance of Machine Learning Algorithms for Heart Disease Prediction: Logistic Regressions Regularized by Elastic Net, SVM, Random Forests, and Neural Networks. *UTSA Journal of Undergraduate Research and Scholarly Works, ,* Volume Volume 8.

Rimal, Y., Sharma, N., Paudel, S., Alsadoon, A., Koirala, M.P. & Gill, S., 2025. Comparative analysis of heart disease prediction using logistic regression, SVM, K-nearest neighbours and random forest with cross-validation for improved accuracy. *Scientific Reports,* Volume 15(1), p. Article 13444.