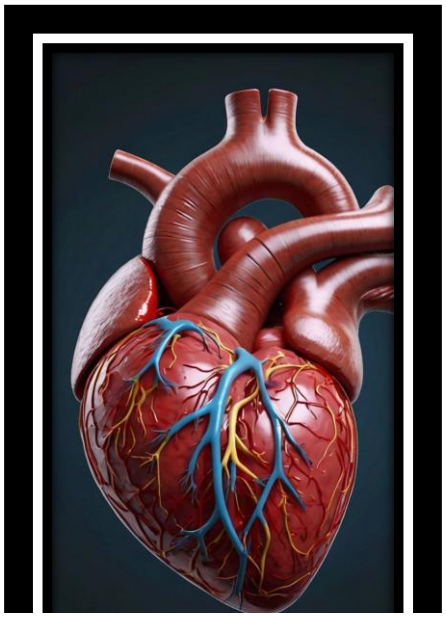

Applying Machine Learning Techniques and Optimization to Diagnose Ischemic Heart Disease (IHD) in Diabetes Patients in Bangladesh: Investigations from a Cross-Sectional Study in 2024

Submitted By Mir khalid Hassan

ID: 20231167 Batch: 11th Section: B

Index

Letter of Endorsement

To Whom It May Concern:

The research project called **"Applying Machine Learning Techniques and Optimization to Diagnose Ischemic Heart Disease (IHD) in Diabetes Patients in Bangladesh: Investigations from a Cross-Sectional Study in 2024"** was turned in by **Mir Khalid Hassan** (ID: 20231167, Batch: 11th, Section: B) and has been approved because it is academically sound. I oversaw the research at **Jahangirnagar University's Department of Statistics and Data Science**.

I have meticulously observed the advancement of this project and have actively contributed guidance at each phase of its evolution. The student has demonstrated exceptional commitment and diligence in conducting this study. This letter constitutes formal acceptance of the research, and I encourage its submission as a requisite for the course.

Please don't hesitate to contact me if you need more explanation or information.

Respectfully,

Md. Habibur Rahman,
Associate Professor, Department of Statistics and Data Science,
Jahangirnagar University, Savar, Dhaka-1342, Bangladesh.
Cell Phone: +8801942246023.
Email: habib.drj@juniv.edu, habib.drj@gmail.com.

Recognition

I wish to convey my deep appreciation to my respected supervisor, **Md. Habibur Rahman, Associate Professor, Department of Statistics and Data Science, Jahangirnagar University**, for his outstanding supervision, steadfast support, and essential insights during this project. His proficiency and thorough methodology have been important in developing this work and achieving its completion.

I am profoundly grateful to the faculty and staff of the **Department of Statistics and Data Science** at **Jahangirnagar University** for supplying essential resources and for their unwavering support throughout my research.

I would like to express my profound gratitude to all participants in this study, whose collaboration and engagement have been essential to the successful completion of this project.

Finally, I express my gratitude to my family and friends for their unwavering encouragement, patience, and understanding, which have served as a continual source of motivation during my academic pursuits.

Submitted By:

Mir Khalid Hassan
ID: 20231167
Batch: 11th
Section: B

Abstract

Background: Ischemic heart disease (IHD) is a predominant cause of morbidity and mortality, especially among diabetic individuals in Bangladesh. Early detection of ischemic heart disease is essential for appropriate intervention and improved patient outcomes. **Machine learning (ML)** methodologies have demonstrated potential in improving diagnostic precision for ischemic heart disease (IHD). This project is to utilize machine learning methodologies and optimization techniques to enhance the identification of ischemic heart disease in diabetic patients in Bangladesh. The main goal of this research is to examine the efficacy of machine learning techniques, encompassing algorithm optimization and ensemble learning, for identifying ischemic heart disease in diabetic patients via a cross-sectional study done in 2024.

Methods: A dataset comprising clinical, demographic, and laboratory data from diabetic patients was analyzed utilizing fourteen distinct machine learning algorithms: **Logistic Regression (LR), k-Nearest Neighbors (kNN), Naive Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), Ridge Classifier (RC), Random Forest (RF), Quadratic Discriminant Analysis (QDA), AdaBoost, Gradient Boosting (GB), Linear Discriminant Analysis (LDA), Extra Trees Classifier (ETC), Classifier Chain (CC), and Decision Forest (DF)**. **Five-fold cross-validation** was utilized for **hyperparameter adjustment** to enhance the model's predictive accuracy. The effectiveness of various models was evaluated based on accuracy and precision, with a specific focus on assessing their generalization abilities using accuracy analysis for each fold.

Results: The evaluation metrics indicate that Gradient Boosting is the most effective model, with an accuracy of 0.910 and a ROC AUC of 0.9694. It consistently outperforms other models in these areas, indicating that it is the most reliable model for classification tasks where predicted accuracy and class distinction are critical. However, in terms of log loss, Random Forest demonstrates higher performance with a lower value of 0.2493, signifying its enhanced reliability in probabilistic predictions. Nevertheless, Gradient Boosting's exceptional performance in accuracy and ROC AUC makes it the preferable choice for most applications.

Conclusions: The study highlights the potential of **Machine Learning Techniques** and Optimization, with Gradient Boosting being the most effective model for overall classification performance. Random Forest is a viable alternative for reducing log loss. Soft Voting Ensemble, while successful, falls short of the top models. The study also suggests that integrating GridSearchCV, five-fold cross-validation, and soft voting ensemble classifiers can help in early diagnosis and treatment planning for diabetes patients with ischemic heart disease.