



ICHSECMICE -2025
11-12th October 2025

Sardar Patel Institute of Higher Education, Kurukshetra

Prediction of Heart Disease Using Data Mining and Machine Learning: A Case Study of Patna

Jyoti, Department of Computer Science & Engineering, Sandip University, Sijoul
Dr. Prena Nagpal, Department of Computer Science & Engineering, Sandip University, Sijoul

Abstract

Cardiovascular diseases are a leading cause of mortality worldwide, and India faces a growing burden due to rapid urbanization and changing lifestyles. Most current predictive models for heart disease rely heavily on generalized datasets such as the UCI repository, which may not accurately capture the demographic and clinical realities of regional populations like those in Patna, Bihar. This research integrates primary data from local institutions—including NMCH and IGIMS—supplemented with private healthcare providers and government statistics, with the UCI Heart Disease dataset. Key clinical and demographic variables include age, sex, blood pressure, cholesterol, and fasting blood sugar. Modern machine learning algorithms—Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN)—are implemented using Python's scikit-learn and Tensor Flow/Keras libraries. Model performance is evaluated by ROC-AUC, with interpretability assessed using SHAP values. Results demonstrate that models trained on regionally augmented data outperform those relying solely on global datasets, with Random Forest yielding an ROC-AUC of 0.89 and SHAP analyses highlighting blood pressure and cholesterol as the most significant predictors. The study underscores the importance of regional data integration and interpretable machine learning for more actionable and equitable healthcare in Patna.

Keywords: Heart Disease, Machine Learning, Random Forest, SVM, ANN, Python, SHAP, ROC-AUC, Patna, UCI, NFHS-5