

Disease Prediction System using Machine Learning

Early detection of chronic diseases such as diabetes, cardiovascular disease, and stroke significantly improves patient outcomes, yet preliminary screening remains inaccessible in many settings. This project, MediPredict, implements a machine-learning–based disease prediction system that estimates individual disease risk from structured clinical input features and presents results through an intuitive symptom-checker web application.

The system is built on publicly available clinical datasets — the diabetes dataset, the UCI Heart Disease dataset, and a stroke prediction dataset. Preprocessing includes missing-value imputation, categorical encoding, feature scaling, and stratified train–test splitting. A comparative modelling framework trains multiple classifiers, including Logistic Regression, Support Vector Machines, Random Forest, Gradient Boosting, and k-Nearest Neighbours, evaluated with stratified k-fold cross-validation using accuracy, precision, recall, F1-score, and ROC-AUC as metrics. The best-performing model per disease is serialized and served by a Flask/Streamlit application where users enter clinical parameters and receive an instant risk probability, severity band, and explanation of the prediction via feature importance and SHAP values.

Evaluation demonstrates strong classification performance across all three disease targets, with cross-validated metrics documented in the report. The application layer supports multi-disease screening, batch predictions on CSV input, and an evaluator dashboard showing confusion matrices and ROC curves.

Limitations include reliance on public datasets that may not represent all populations and the probabilistic nature of predictions. MediPredict is positioned as an educational screening and demonstration tool for final-year projects, not a clinical diagnostic device; the application carries a medical disclaimer accordingly. The complete project package includes full source code, trained model artifacts, a formatted project report, a viva presentation, and documentation, making it suitable for final-year B.Tech/M.Tech submissions and academic demonstration.