A Hybrid PF-PSO and LGBM-RFE Feature Selection Framework for Chronic Disease Prediction Using Ensemble Learning

Salwani Abdullah
Centre for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Malaysia

Lan Gongwei
Centre for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Malaysia

Zalinda Othman
Centre for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Malaysia

Abstract

Chronic conditions, specifically diabetes and cardiovascular diseases, remain leading drivers of global mortality, demanding highly accurate and timely predictive tools. However, the efficacy of clinical machine learning models is frequently compromised by high-dimensional datasets containing redundant features and highly non-linear variables. To address these challenges, this study introduces a novel hybrid feature selection framework that synergistically integrates Perturbation Factor–Particle Swarm Optimization (PF-PSO) with Light Gradient Boosting Machine Recursive Feature Elimination (LGBM-RFE). This dual-stage methodology achieves an optimal balance between global exploration (metaheuristic optimization) and local exploitation (algorithmic elimination) to isolate highly informative clinical features while discarding noise. The optimized feature subsets were rigorously evaluated across four predictive models: Random Forest, Logistic Regression, Support Vector Machine, and an ensemble classifier. Benchmark evaluation demonstrates that our proposed framework systematically reduces feature dimensionality while consistently maintaining or enhancing classification performance. Notably, the ensemble classifier achieved superior accuracy, reaching 0.7922 for diabetes prediction and 0.9180 for heart disease prediction, consistently outperforming all individual baseline models. These results demonstrate the robustness, scalability, and interpretability of the proposed hybrid framework, establishing it as a highly viable clinical decision-support tool for early disease risk detection and intervention.

Keywords

Chronic Disease Prediction, Hybrid Feature selection, Particle Swarm Optimization, LightGBM Recursive Feature Elimination (LGBM-RFE), Ensemble Learning.