

Comments and Suggestions for Authors

The paper addresses a relevant problem at the intersection of machine learning and public health, and I appreciate the authors' emphasis on interpretability over mere predictive accuracy. That said, several aspects warrant deeper consideration. The choice of models — Logistic Regression, Decision Tree, Random Forest, and a basic feed-forward Neural Network — is rather conservative. Given that the study positions itself within the explainable AI space, the omission of gradient boosting methods such as XGBoost or LightGBM, which pair naturally with SHAP-based explanations and routinely outperform the selected models on tabular clinical data, is a notable gap. Furthermore, SHAP itself is conspicuously absent from the XAI toolkit despite being the most widely adopted and theoretically grounded framework for model-agnostic explanation; relying solely on PDP, ICE, and LIME leaves the interpretability analysis incomplete, particularly at the global level.

The multi-centre dataset is a commendable aspect of this work, yet a sample of only 100 patients across multiple hospitals raises serious concerns about statistical power and the reliability of cross-validated performance estimates. With such a small cohort, ten-fold cross-validation folds contain very few samples, and the decision to forgo hyperparameter tuning — while understandable given the interpretability focus — makes it difficult for readers to judge whether the models have been given a fair chance before their explanations are scrutinised. The paper would benefit from reporting standard classification metrics such as accuracy, F1-score, and AUC alongside the explainability plots to allow the community to contextualise the explanations against actual model competence.

From an EEE-to-clinical translation standpoint, I would encourage the authors to discuss how this framework could be embedded within a real-time clinical decision support system — for instance, as an edge-computing module in hospital monitoring infrastructure. Addressing latency, computational constraints, and how clinicians would interact with LIME outputs at the bedside would significantly strengthen the paper's practical contribution and better bridge the engineering and healthcare domains.