Brain Stroke Prediction using Gradient Boosting with XAI

Brain stroke is a major global health issue, being the second leading cause of death and disability worldwide (WHO-2023). It happens when blood flow to part of the brain is blocked or reduced, depriving brain cells of oxygen and nutrients. This can cause brain cell death and lead to long-term difficulties with movement, speech, and daily activities. The brain relies on a continuous blood supply for proper function, and a brain stroke disrupts this vital process.

Ushasree D et al. have made a significant contribution to brain stroke prediction with their Enhanced Stroke Prediction Ensemble using Stacking Methodology (ESPESM). This framework combines a hybrid feature engineering approach with a stacking methodology. ESPESM employs K-Nearest Neighbors, Random Forest, Gaussian Naïve Bayes, and Support Vector Machines as base learners, with Random Forest serving as the meta-learner to refine predictions. However, it has some limitations: it focuses on traditional machine learning algorithms for brain stroke prediction, which may not fully capture the potential improvements offered by more advanced techniques. Additionally, it does not investigate methods beyond SMOTE for addressing imbalanced datasets and does not incorporate emerging technologies like Explainable AI (XAI), which could further enhance both interpretability and accuracy in brain stroke predictions.

The main objective of this research is to introduce a new framework to enhance brain stroke prediction by integrating gradient-boosting algorithms with Explainable AI (XAI) techniques. The proposed approach will utilize a comprehensive dataset that includes important risk factors for brain stroke and requires careful preprocessing to handle missing values and encode categorical variables. To address class imbalances, a hybrid data balancing approach combining Synthetic Minority Over-sampling Technique (SMOTE) and Random Under sampling will be utilized. Various boosting algorithms, including Gradient Boosting, XGBoost, LightGBM, and CatBoost, are optimized via hyperparameter tuning to improve predictive accuracy. XAI techniques such as Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) are integrated to boost model interpretability, providing insights into individual predictions and overall model behavior. This research aims to enhance prediction accuracy and offer a clearer understanding of brain stroke risk factors, potentially leading to more effective early interventions and better patient outcomes.

REFERENCES

1. Ushasree, D., Praveen Krishna, A.V., & Mallikarjuna Rao, C. (2024). Enhanced stroke prediction using stacking methodology (ESPESM) in intelligent sensors for aiding preemptive clinical diagnosis of brain stroke. Measurement: Sensors, 33, 101108.
2. K. Mridha, S. Ghimire, J. Shin, A. Aran, M. M. Uddin, & M. F. Mridha (2023). Automated stroke prediction using machine learning: An explainable and exploratory study with a web application for early intervention. IEEE Access, 11, 52288-52308.
3. World Health Organization. (2023). The top 10 causes of death. Retrieved from https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death.