Paper No: SC-25 **Smart Computing**

An Ensemble Machine Learning Approach for Stroke Prediction

P. Premisha*
Dept. of Information Communication Technology
University of Vavuniya, SriLanka
premishaprem@vau.ac.lk

Senthan Prasanth
Department of Physical Sciences and Technology
Sabaragamuwa University of Sri Lanka
sprasanth@appsc.sab.ac.lk

Mauran Kanagarathnam
Department of Physical Science
University of Vavuniya, Sri Lanka
maurank@vau.ac.lk

Kuhaneswaran Banujan
Department of Computing and Information Systems
Sabaragamuwa University of Sri Lanka
bhakuha@appsc.sab.ac.lk

Abstract **- Nowadays, one out of four people above 25 will suffer from a stroke. Especially this year, with the highest countof around 13.7 million people discovered with stroke for the firsttime. Out of 13.7 million, 5.5 million were fatalities. This was stated in a recent WHO study. It is estimated that if no action istaken, the number of fatalities will rise to 6.7 million yearly. Thepandemic situation of COVID-19 will play a significant cause inthe expanded death rate of stroke. Even for adults and patientswith minor risk factors affected by stroke rather than in previous years. This study predicts the impact level of stroke with the development of an ensemble model by combining the various classifiers performed well in isolation. Predicting the stroke status in patients would help the physicians determine theprognosis and assist them in providing the targeted therapy in alimited time. During this study, an ensemble model was built byconsidering the base, bagging, and boosting classifiers: SupportVector Machine, Naïve Bayes, Decision Tree, Logistic Regression, Artificial Neural Network, Random Forest, XGBoost, LightGBM, and CatBoost. The dataset consists of 5110 patient details, along with 12 attributes that were analyzedin this research. The final ensemble model was developed by carrying out the methodology in two phases. During the first andsecond phases, the classifiers mentioned above were trained without hyper-parameter tuning and with hyperparameter tuning and tested against the fundamental evaluation matrices.During each phase, the classifier that produces the highest classification accuracy is discovered from the base, bagging, andboosting categories. From the results obtained, the final ensemble model was constructed using the Max Votingapproach, which yielded an accuracy of 95.76%.**

Keywords - bagging and boosting, ensemble, Machine Learning, medical informatics, stroke