CHAPTER 9

CONCLUSION AND FUTURE ENHANCEMENTS

9.1 CONCLUSION

Silent heart attacks, a type of cardiovascular disease, have an extremely low survival rate and must be discovered at an extremely early stage in order to treat the underlying cause of the disease effectively. An effective heart disease prediction system is necessary because of the rise in cardiac patients caused by sedentary lifestyles and a shortage of medical specialists. There are a wide variety of ML methods available that make use of both stand-alone classifiers and hybrids. However, these models' results varied considerably between different cardiac datasets or they were not modeled using low and high dimensional data. Also, as the size of datasets used for predictive modeling grows, feature selection and prediction become more feasible. Existing feature selection and prediction approaches still struggle with datasets that have complicated feature relationships and significant degrees of redundancy. Therefore, this study effectively identifies the issues with existing feature selection methods and classification methods and proposes a new model for an improved heart attack prediction infrastructure.

The first stage of work involves feature selection by Feature Importance (FI) score of Gradient Boosting algorithms, where feature subsets are formed based on the FI Score as a threshold beginning with the top rank. By pinpointing the threshold range in whichthe optimal feature subset can be obtained, this work suggests a substantial decrease in the search space of feature subsets. When compared to an exhaustive search of relevant feature subsets of the heart datasets, there would be a saving in the number of searches done.

The next step proposes an algorithm for feature selection called MBAR to pick out the most important characteristics for making heart disease forecasts. CatBoost (CatB) is used as the base model in this approach, and a novel feature elimination technique is employed. A comparison of MBAR and existing BoostARoota (BAR) algorithm is done and it is inferred that MBAR needs lesser computation time than its counterpart. MBAR with classification by CatB outperformed the other feature selection algorithms which were experimented on the heart datasets.

After features have been chosen using MBAR, in the second stage, new research proposes using a Super Learner Ensemble Model (SLEM) to analyze the data. A number of different ML base models are combined to form the SLEM model. These models include SVM, LR, KNN, GNB, RF, DT, MVE, XGB, and CatB. These models were chosen by three rounds of stratified cross validation with 10 folds. The predictions of these classifiers were input to the meta learner LR. These classifiers were trained on the entire dataset also. Through a process of backward elimination, they discovered that combining the SLEM classifiers CatB and DT improved both classification time complexity and performance. When applied to features chosen by MBAR, the SLEM model performed better than when applied to unselected features or to datasets with no feature selection at all.

The third stage of the research process involves the creation of a novel OSLEM, where the optimal classifiers in the SLEM ensemble model are chosen using the Whale Optimization Algorithm (WOA). In ensemble pruning, we introduce the pairwise divergence measure to consider both the combined classifiers' accuracy and the pairwise diversity among those classifiers. The WOA method uses a population of whales to store information about the number of base classifiers used and whether or not they made it into the final ensemble. The whale population is then subjected to WOA, at which point a fitter group of whales is produced. The next stage is to calculate each whale's score on the diversity index and then pick the one with the highest score. The best-performing whale is then used to make predictions about the test data instances for the heart disease prediction, and SLEM is employed to learn the predicted class label.

Accuracy, Precision, Specificity, Sensitivity, and F1-Score on high and low dimensions' heart datasets are compared for the proposed OSLEM classifier that is modeled on the features identified by MBAR. The study shows that the MBAR with OSLEM is more efficient than other existing models across the heart datasets used in the study. The suggested model is trustworthy for predicting cardiovascular illness due to its excellent levels of precision, recall, f1-score, and accuracy. Other highlights of the proposed model are:

- In order to enhance generalizability and reduce bias, this research work has applied the proposed model on datasets of diverse population, namely USA, South Africa and India.
- As the imbalanced datasets used in this work are balanced by applying SMOTE before training the proposed model, the class imbalance problem which can lead to unfair models or inaccurate predictions is avoided.
- ML models are inherently interpretable because their structure allows easy tracing of decisions. This interpretability is crucial in healthcare for clinician trust and regulatory approval.

Through improved data quality, explainable AI techniques, rigorous validation, and ethical considerations promising research can be translated into impactful clinical applications.

9.2 FUTURE WORK

ML models are more interpretable and computationally less expensive and works well with small, structured datasets. Whereas DL models often considered as "black boxes" due to complex architectures, require high-performance GPUs and works well with unstructured datasets. By coalescing information from various sources including health records, medical imaging, human activity monitoring devices, and genomic data and integrating Machine Learning (ML) and Deep Learning (DL) models, for heart disease prediction, researchers can gain deeper insights into the factors contributing to heart disease and develop more comprehensive prediction models that would enhance interpretability and prediction accuracy.

Exploration of several other optimization algorithms would contribute to the development of efficient and effective model to predict the risk of heart diseases.

By applying ML techniques, researchers can develop personalized risk assessment models. This approach can help optimize treatment plans and interventions, improving patient outcomes and reducing healthcare costs. These models can help healthcare professionals make better-informed decisions and allocate resources more efficiently.

Cardiac Biomarkers in blood will raise during a heart attack. Medical practitioners rely on these biomarkers to diagnose heart attack in the cases where there is no significant variation in the ECG. Research experimenting cardiac biomarkers data of persons having family members with heart disease, might be a new pathway leading to early forecasting of heart disease.

Explainable Artificial Intelligence (XAI) techniques must be used to improve transparency in ML predictions. This would help clinicians understand the reasoning behind a risk assessment and thereby build trust and make informed decisions.

Overall, there is a lot of potential for improving early detection, tailored therapy, and clinical decision support through ML research in the prediction of cardiac disease. Researchers hope to make great achievements against heart disease, as well as improve patient outcomes and lower healthcare costs, by harnessing the power of ML.