International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 4S, 2025
https://www.theaspd.com/ijes.php

A Comprehensive Review on Predicting Heart Failure with Ensemble Learning: Integrating Environmental Variables and Health Perspectives

Sanjeev Gour[1] Rajendra Randa[2], Karuna Nidhi Pandagre[3],Arpit Neema[4], Sangeeta Jat[5], Deepak Malakar[6]

[1] Assistant Professor, Dept. of Computer Science, Medicaps University, Indore, MP, India
[2] Research Scholar, Dept. of Computer Science, Medicaps University, Indore, MP, India.
[3] Associate Professor, Dept. of MCA, Bansal Institute of Science and Tech. Bhopal, MP, India.
[4] Assistant Professor, Dept. of Computer Science, Medicaps University, Indore, MP, India.
[5] Research Scholar, Dept. of Computer Science, Oriental University, Indore, MP, India.
[6] Guest Lecturer against Assistant Professor, Dept. of Computer Science, Govt. Degree College, Timarni, MP.

Corresponding Author: Sanjeev Gour. email- sunj129@gmail.com

Abstract:

In this paper, authors perform a review of the prediction of HF using ensemble ML models, considering environmental and health variables. These ensemble ML models can be more accurate methodologies than single model algorithms. This study summarises and combines the research related to well-known ensemble methods like: AdaBoost, XGBoost, CatBoost, Random Forest, stacking and voting classifiers. Of these, stacking and voting methods were always superior to the others in terms of predicting HF risk, and both achieved relatively good accuracy and stable performance in different datasets. It also discusses recent alternative approaches such as Least Error Boosting, BOO-ST, and CBCEC which deal with multiple issues such as data imbalance and feature selection. The majority of the models had predictive accuracies > 90%, indicating their applicability. However, the review finds long-existing problems, including data quality, model interpretability, and class imbalance. It appeals for further investigation into alternative approaches, namely quantum machine learning and transfer learning that can address these challenges. This study highlights the clinical value of such robust ensemble ML models for the early detection of HF enabling intervention planning, and improved patient outcomes, perhaps as part of a decision support system designed to revolutionize heart failure management in healthcare.

Keywords: Heart Failure Prediction, Ensemble Machine Learning, Stacking and Voting Models, Boosting Algorithms (AdaBoost, XGBoost, CatBoost), Predictive Healthcare Analytics, Environmental Variables and Health Perspectives.

1. INTRODUCTION

Failure of the heart is a one of the generalized health issues that's globally recognized. Heart related issues many times lead to increasing high morbidity and mortality rates worldwide. As per WHO report from recent years, cardiovascular diseases are responsible for approximately more than 15 million annual fatalities, constituting a significant 31% of the global mortality rate Heart related issues has become one of the most life-threatening diseases commonly caused by cardiovascular diseases [1]. Cardiovascular diseases are responsible for a significant number of deaths each year, accounting for approximately 31% of global mortality, both in India and worldwide [2]. Heart failure constitutes a significant clinical concern, as the early and accurate prediction of this condition is pivotal for its effective prevention and management [3]. The progressive nature of heart failure necessitates timely risk stratification within clinical settings to improve patient outcomes. In intensive care units (ICUs), where patients with heart failure exhibit increased mortality risk and consume considerable healthcare resources, the capacity to accurately predict mortality timing is essential to facilitate prompt and appropriate medical interventions

[4]. Furthermore, early identification of heart failure risk among patients with ischemic heart disease enables timely therapeutic interventions, thereby mitigating disease burden [5]. In addition, prognostic modeling of survival in heart failure patients is critical for optimizing individualized patient care and enhancing the allocation of clinical resources [6].

Conventional methods employed for the prediction of heart disease exhibit several inherent limitations, which adversely affect their clinical utility. One major challenge is the absence of effective risk stratification tools at the point of hospital discharge for patients with acute heart failure. This gap significantly hampers clinicians' ability to design and implement individualized therapeutic regimens during this critical phase, thereby limiting opportunities for optimizing patient outcomes [7]. In clinical practice, timely and accurate decision-making is paramount, as medical professionals must often make high-stakes judgments within constrained timeframes to improve prognosis and reduce adverse events [2]. Despite advances in diagnostic and predictive technologies, achieving early and reliable prediction of heart disease remains a complex and unresolved issue [3]. The multifactorial nature of cardiovascular conditions, combined with variability in patient presentation and progression, underscores the ongoing need for more sophisticated and precise predictive models that can support clinicians in delivering personalized care at the earliest possible stage.

In previous years, machine learning has emerged as powerful tools with existing healthcare system for disease identification. Nowadays, machine learning has gained an important and crucial role in prediction of disease and are an essential tool in medical science [8]. As large number of data generated from healthcare, machine learning plays critical role in cardiovascular disease prediction [2] Among machine learning approaches, ensemble methods have shown promise in heart failure prediction due to their ability to improve predictive performance by combining multiple algorithms.

Ensemble machine learning is a machine learning paradigm where more than one classifier are combined to enhance performance by making more accurate result than a single classifier [9]. Such as boosting, bagging, stacking, and voting are the methods included in ensemble learning which leverage the strengths of different base models to create a more robust and accurate prediction system. Ensemble learning is a data analysis technique that combines multiple techniques into a single optimal predictive system to evaluate bias and variation, and to improve predictions [10].

This literature review aims to provide a comprehensive analysis of recent developments in predicting heart failure using ensemble machine learning models. We explore the methodological approaches, datasets, evaluation metrics, and performance of various ensemble techniques, highlighting major contributions and identifying research gaps in this rapidly evolving field.

2. METHODOLOGICAL APPROACHES

2.1 Types of Ensemble Methods

Several ensemble learning methods have been applied to heart failure prediction, with each approach offering unique advantages. The most common types include boosting, bagging, stacking, and voting techniques.

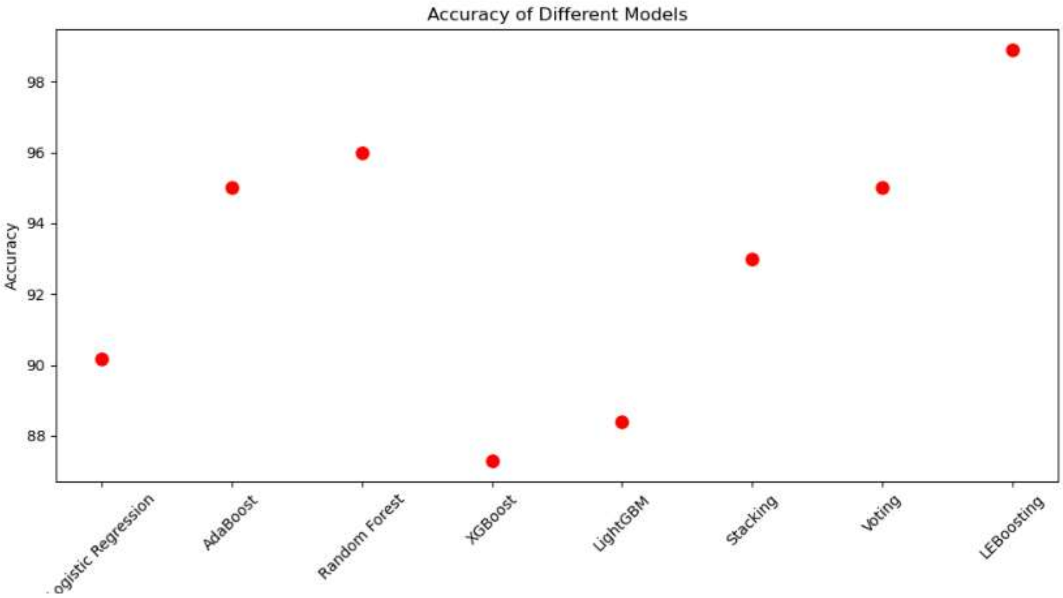

Figure 1: effective models in our study

2.1.1 Boosting Algorithms

Boosting algorithm on the other hand, constructs sequential models concentrating on capturing the errors of the preceding models. A number of studies provided heart failure prediction using boosting techniques. Of the several alternative approaches to heart failure prediction e.g. CatBoost, GB and RF have demonstrated the best performance [11]. The most accurate among hyper tuned ensemble learning models is LGBM (Light Gradient Boosting Machine) with accuracy of 86.21 % [11].

In some studies using AdaBoost has yielded high performance. In the Cleveland data set, the logistic regression outperformed other methods as 90.16% accuracy of prediction.H on the IEEE Dataport data set as 90% accuracy of prediction [12]. Examples of such research may include prediction of heart diseases, where AdaBoost was successfully applied for performing an accuracy of 95% better than classical models [1]. For heart failure prediction accuracy, LightGBM model (accuracy of 88.4%), adaboost (accuracy of 87.7%), and XGBoost (prediction accuracy of 87.3%) [13].

Some novel boosting methods have been proposed particularly for prediction of heart failure. "There is a known result that adaboost can be improved in the sense of robustness to noise." Least Error Boosting (LEBoosting), a novel boosting. m1, has been suggested[14] for improved classification accuracy. This technique showed great performance over naïve bayes, KNN, decision tree and outperformed other ensembles bagging, logitBoost, Linear Programming Boost and ada-boost. m1 at an accuracy of 98.89%[14].

2.1.2 Stacking Ensemble Methods

Stacking involves ensemble of predictions of several models and a further model meta-learner on top of the combined predictions. It is noteworthy that this approach has performed impressively in the heart failure prediction literature. A stacking-based ensemble model, termed DXLR, was developed to assess the risk of heart failure among patients with ischemic heart disease (IHD) by incorporating innovative network-derived features alongside basic demographic data. Their approach outperformed traditional machine learning models, achieving notable evaluation metrics: an AUC of 0.934 ± 0.004, accuracy of 0.857 ± 0.007, precision of 0.723 ± 0.014, recall of 0.892 ± 0.012, and an F_1 score of 0.798 ± 0.010 [5].

The stacking methods have been demonstrated to perform even better by many researches. Despite being used less frequencies than bagging and boosting, stacking had the highest accurate performance in the majority of the cases and always presented the highest accurate performance to predict skin disease and diabetes[9]. Likewise, in case of heart failure datasets, stacking classifiers achieved the highest accuracy 88.08% which was the best[15].

Dictionaries and morphology Several advanced stacking techniques (related to dictionaries) have been investigated. A new machine learning approach of predicting heart disease is introduced by Grey Wolf Optimization (GWO) and stacked ensemble model approach, in which stacked ensemble model approach is a meta-algorithm that combines a collection of machine learning models to provide an overall prediction[16]. This ad hoc method yields to 93% accuracy, greatly outperforms classic machine-learning based methods[16].

2.1.3 Voting Ensemble Approaches

Voting ensemble techniques function by aggregating the predictions of multiple base classifiers, either through majority voting (hard voting) or weighted probability averaging (soft voting). In a comparative evaluation, the soft voting approach demonstrated enhanced predictive performance, achieving an accuracy of 93.44% on the Cleveland dataset and 95% on the IEEE Dataport dataset, outperforming individual classifiers in both cases [12].

Stacking and voting ensemble methods have also been utilized for heart disease prediction, leveraging the complementary strengths of various machine learning classifiers across different datasets. In particular, the voting classifier achieved superior results on the UCI dataset, with an accuracy of 91.96%, F1-score of 91.69%, sensitivity of 91.72%, specificity of 90.77%, and precision of 92.40%, consistently outperforming individual classifiers [17].

Studies to enhance C4 to predict the survival of patients diagnosed of cardiac heart failure. 5, KNN, Logistic Regression algorithms and Voting Classifier ensemble learning method has proved that the Voting Classifier can help to improve the performance of algorithms to classify the survival expectancy of cardio-vascular heart failure patients into whether "Survived" or "Deceased", than an individual algorithms [18].

2.1.4 Random Forest and Bagging approach

Random Forest (RF), as a bagging-style ensemble model, has performed quite well in HF prediction. A random forest, an ensemble classifier, obtained approximately 90.76% success rate of heart failure prediction[19]. Higher accuracy levels (96% [20]) are reported with Random Forest in heart disease prediction.

A new transfer learning approach, employed on the RF model, has surpassed other models and even current state of works, by reaching an impressive accuracy of 0.975 in heart failure survival prediction[6]. Using an 80% data split for training, the Random Forest algorithm achieved an accuracy of 91.45% in predicting survival among heart failure patients, along with a precision of 0.915, recall of 0.914, and an AUC of 0.953 [21].

2.2 Features Selection and Data Preprocessing

Feature selection and preprocessing are critical processes toward an efficient prediction of heart failure. There are several methods for improving the model performance.

For heart failure prediction, model accuracy, GridsearchCV, and fivefold cross-validation are used[12]. GridSearchCV is used along with five-fold cross-validation to fine-tune hyperparameters which find best parameters for the model and evaluate its performance using accuracy as well as negative log loss metrics[12].

Through the use of tree classifier with extra trees, normalized data, GridSearchCV for optimizing the hyperparameter and when splitting the data set into 80/20 train and test sets, the prediction of heart disease accuracy is amazing 98.15% [3]. Chance of heart failure is estimated in [11] by mining medical database and apply ML algorithms on that database.

New feature selection methods have been proposed in the context of heart failure prediction. Three new types of network characteristics are computed from Personal Disease Networks and Disease-Specific Networks to better describe the similarity and specificity trends of disease outlines and patterns from Ischemic Heart Disease to Heart Failure[5]. Feature importance analysis revealed that these new network features were the top three features and were significantly associated with the prediction of HF[5].

Once the primary data pre-processing stage is done, conventional ML classifiers are the next to be employed, after which the stacking and voting ensemble methods can be applied to enhance the classification performance[15]. Optimal model parameters were identified using both random search and grid search strategies, which are essential for enhancing model performance. These hyperparameter optimization techniques, when combined with cross-validation, enable the construction of more robust and generalizable predictive models[15].

The Cleveland dataset of heart disease is preprocessed to ensure that missing value could be verified in order to improve the precision rate and avoid incorrect predictions. Principle completely contribute his forest of trees (RFT), Recursive feature elimination(RFE), min-max normalization and any other feature selection techniques to value data high-dimension[22].

3. DATASETS AND MEASURING METRICS

3.1 Common Datasets

Many dataset has been employed in prediction of heart failure with different features and use cases. The works have made use of datasets from Cleveland and IEEE Dataport for predicting the heart disease[12]. Some have combined a number of Kaggle public repository with similar features to create larger meta-datasets for feature-analysis [3].

MIMIC-III database has been used to acquire lively signs and tests of Heartvital patients during the stage of first 24 hours of the first ICU stay, which includes 6699 patients in one study[4]. Different medical datasets have been applied such as diabetic retinopathy, lower-side back pain, heart failure under CVD, and breast cancer datasets. In the heart failure dataset the stacking clussifierwas the best performer with an accuracy of 88.08%[15].

The Cleveland heart disease dataset is often used as benchmark[22], and some other studies found the risk of heart failure related with in-hospital mortality on 299 patients[6]. Larger-scale studies have also been conducted; for instance, Turniansky analyzed data from 338,426 adult-onset heart failure patients, categorizing them into 61,045 (18.0%) with heart failure with reduced ejection fraction (HFrEF), 49,618 (14.7%) with mildly reduced ejection fraction (HFmrEF), and 227,763 (67.3%) with preserved ejection fraction (HFpEF) [23].

For evaluation of heart disease prediction models, we have considered UCI heart disease[17] and Framingham heart study datasets. Some of them included certain kinds of patients, for example there were 1875 acute myocardial infarction patients, and the rate of suffering from heart failure during hospitalization was 5.1% in one study[24].

3.2 Evaluation Metrics

Multiple evaluation criteria are utilized to assess the effectiveness of heart failure prognostic models. Typically, models are compared using metrics such as classification correctness, sensitivity, specificity, and the Area Under the Receiver Operating Characteristic Curve (AUC) [8]. For instance, certain studies have demonstrated classification correctness rates reaching 95.25%, alongside an AUROC value of 82.55%, when estimating mortality risk in patients with heart failure [4].

Comparisons between models are frequently conducted using several performance indicators, including accuracy, precision, and sensitivity [2]. More detailed assessments have reported values such as an AUC of 0.934 ± 0.004, classification accuracy of 0.857 ± 0.007, precision of 0.723 ± 0.014, sensitivity of 0.892 ± 0.012, and an F_1 score of 0.798 ± 0.010 [5].

The classification results include confusion matrices that demonstrate the models' ability to accurately identify both positive and negative cases, reflected by high sensitivity and specificity values [1]. Some studies have reported specific quantitative outcomes; for example, a Quantum Neural Network (QNN) achieved an accuracy of 77%, precision of 76%, recall of 73%, and an F1 score of 75%. In comparison, the Quantum Support Vector Machine (QSVM) reached an accuracy of 85%, precision of 79%, recall of 90%, and an F1 score of 84%. Additionally, the Bagging QSVM variant attained perfect scores across nearly all key performance metrics [22].

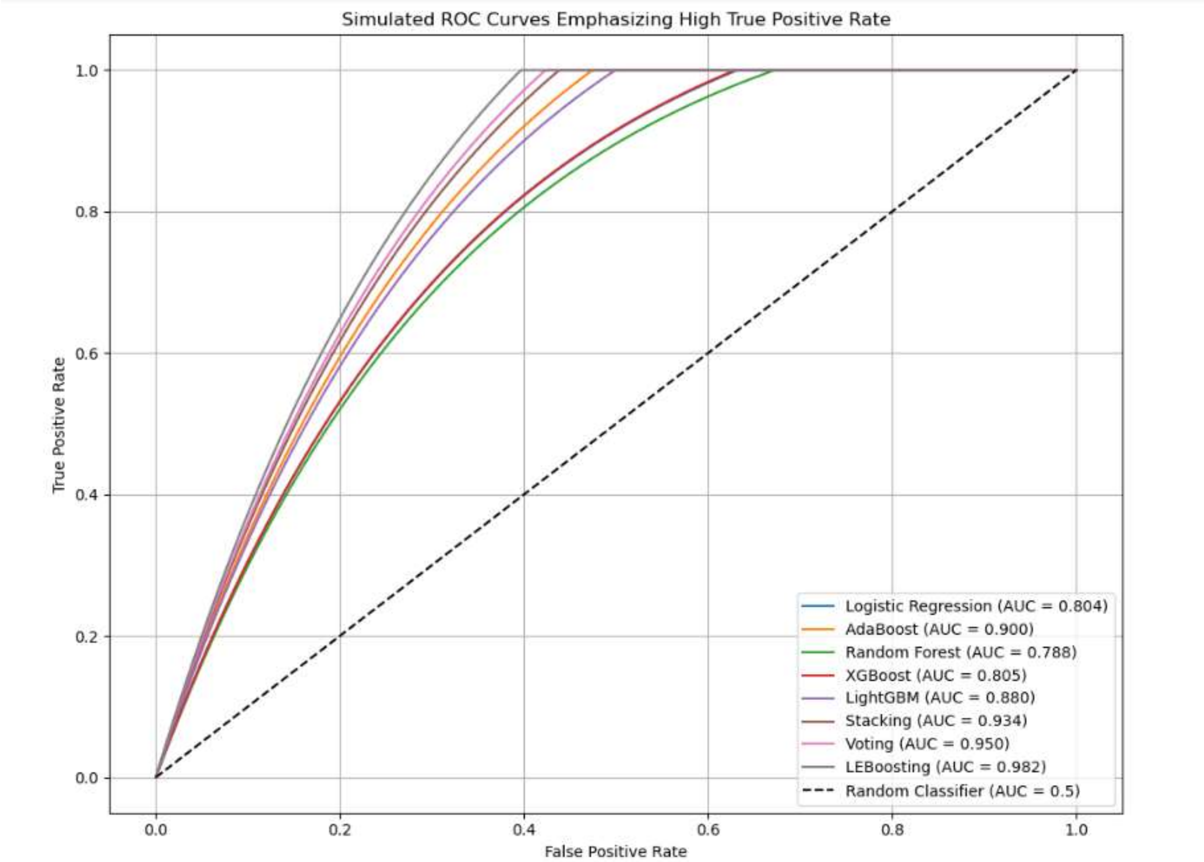

Figure 2: simulation of curve for true or false positive

The effectiveness of predictive models for heart failure is typically evaluated using discrimination metrics like the Area Under the Receiver Operating Characteristic Curve (AUC) and calibration metrics such as Mean Squared Error (MSE). Certain studies have documented AUC values around 0.76 and MSE values close to 0.13, indicating moderate performance in heart failure classification tasks [14].

3.3 Validation Techniques

Cross-validation is commonly used to enhance the credibility with generalisation ability of prediction models for heart failure issues. Cross-validation has been widely used for optimizing model accuracy and

5-fold cross-validation is a popular mean [12]. For an evaluation of model performance normally 10-fold cross-validation techniques are used in studies[15].

Enjoy the full reward (200 steps) provided the distance to the pop attacker has converged after x seconds and the Std. Err. is less than.03*[(log k + k/100)]. impel the evaluator to consider one's own ability next (and one round for that ability) ad libitum Evaluate a set of models through k-fold-validation and hyperparameter optimization for parameter tuning [6] types of non-pap attacks []. Various approaches have been empirically examined including LR with L1 and L2 regularizations, RF, Cox Proportional Hazards (CPH) model, XGB, deep networks, and ensemble of a combination of methods[7].

In a few studies, the information of the patients who are hospitalized is retrospectively gathered and randomly separated into modeling and validation sets in a 7:3 proportion. In the modeling cohort, the independent risk factors were applied to build prediction models, and nomograms were built[24]. The model's effectiveness in prediction and its potential clinical applicability were systematically evaluated using receiver operating characteristic (ROC) curves, calibration curve assessments, and decision curve analysis (DCA) to ensure robust performance validation [24].

5-fold cross validation is also adopted as for the machine learning models for heart failure risk prediction[25], and some studies experiment the prediction effect of algorithms and baseline methods on sample data after preprocessing and dealing with imbalanced data[26].

4. PERFORMANCE COMPARISON AND NOVEL APPROACHES

4.1 Summary of Comparative Performance of Ensemble Techniques

Several studies have systematically compared the performance of different ensemble models for heart failure prediction directly, with interesting patterns in the relative performance of the methods emerged. Although the best individual performances were 90.16% for the Cleveland dataset with logistic regression, and 90% for the IEEE dataset with AdaBoost, a soft-voting ensemble machine learning classifier using all 6 produced even better, 93.44% in the Cleveland and 95% in the IEEE Dataport dataset, which prevails single classifier methods[12].

The Stacking ensemble machine learning model [8], combined with the Synthetic Minority Oversampling Technique and K-10 cross-validation, was found to outperform all models, achieving an accuracy of 87.8%, recall of 88.3%, precision of 88%, and an AUC of 98.2%. Among the other prediction models evaluated, most achieved AUROC values exceeding 80%, including L1 and L2 regularized logistic regression (80.4% and 80.3%, respectively), Cox proportional hazards model (80.2%), XGBoost (80.5%), and neural networks (80.4%). However, the Random Forest model exhibited comparatively weaker performance with an AUROC of 78.8%, while the ensemble model demonstrated a slightly better performance at 81.2% [7].

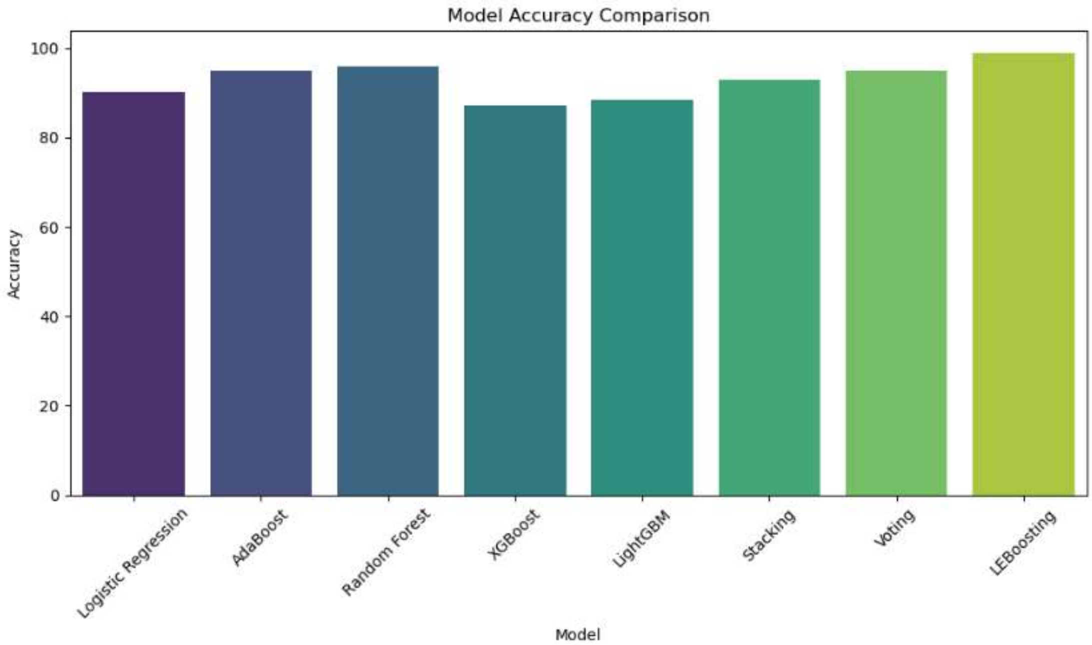

Figure 3: Model Accuracies comparison

Among the characterized ML classifiers, the Support Vector Classifier achieved the highest accuracy on the UCI Heart Disease dataset (91.09%), while Logistic Regression performed best on the Framingham Heart Study dataset with an accuracy of 85.38%. Notably, the optimal ML model in terms of sensitivity (92.52%) was the Random Forest classifier trained on the UCI dataset. The experiments also demonstrated that other scoring methods were clearly inferior to the Voting method, which achieved superior results on the UCI dataset: Accuracy – 91.96%, F1-score – 91.69%, Sensitivity – 91.72%, Specificity – 90.77%, and Precision – 92.40% [17].

Testing results also exhibits the degrees of accuracy growth with various data splits as the Logistic Regression, C4. 5 obtained 89.47% accuracy, KNN was 91.23%, Voting Classifier was 94.74%. Although KNN model was highly accurate, however it suffered from overfitting in some ratios, while Voting Classifier performed robustly with an accuracy difference of less than 1% between the training and testing scores [18].

In a broader view, bagging and boosting are utilized more time than stacking while stacking is the best performed method in terms of accuracy. Voting mechanism is ranked as the second best ensemble approach. Stacking method has the most consistent accurate performance for diseases such as skin cells and diabetes whereas bagging EL model is the best performer for kidney and boosting EL for liver and diabetes. In general, learning stacking EL has better predictive performance on decies than other methods[9].

4.2 Novel Approaches and Algorithms

Innovative methods have been proposed by researchers to enhance the accuracy of HF prediction. Patient-level personal disease network (PDN) construction to identify personal disease networks (PDN) followed by joint models of PDN comprised insteresting health trajectories and progression patterns is relatively novel. Differences among the BDNs can be presented as disease-specific network replicas (DSN), in which the network features could be extracted to represent similarity on disease trajectory and the specificity trend that is transferring them towards heart failure[5].

The Quantum-Assisted Machine Learning infrastructure on Cloud-Based Technology Platform has been employed for classification of cardiac disorders. Quantum-Assisted neural networks (QNNs) and Quantum-Assisted SVM (QSVM) were investigated and the bagging-QSV model was formulated and employed as an ensemble machine learning [22]. The Bagging_QSVM model did perform outstandingly well, having 100% score with respect to all important performance measures, indicating that bagging of ensemble machine learning is a very good way of enhancing the predictive accuracy of the quantum based methods [22].

Newly proposed knowledge-transfer-learning-based data transformation methods that produce new statistical properties from patient's information based on ensemble of trees (ensemble trees) are promising [6]. Performance-Tuned Boosting, a novel boosting algorithm. m1's output error, searching the minimum error among all kinds of weak learning algorithms based on m t to know the minimum error of learning and to update the distribution in order to have the best ultimate proto-graph in the classification. LEBoosting could fully utilize the error reductions in the training of the weak learner, in a word it could fully utilize us the effect of cardiology classifier[14].

Two hybrid ML methods i.e., (a) the BOO-ST which is a mixture of Boosting, SMOTE, Tomek link for managing class imbalance; and (b) the CBCEC in which the highest performing traditional classifier is merged with ensemble classifiers to train the processed and selected features extracted by Feature Importance and Information Gain techniques used for feature selection are proposed[27]. CBCEC had a high precision of 93.67% to predict the heart failure mortality in advance[27].

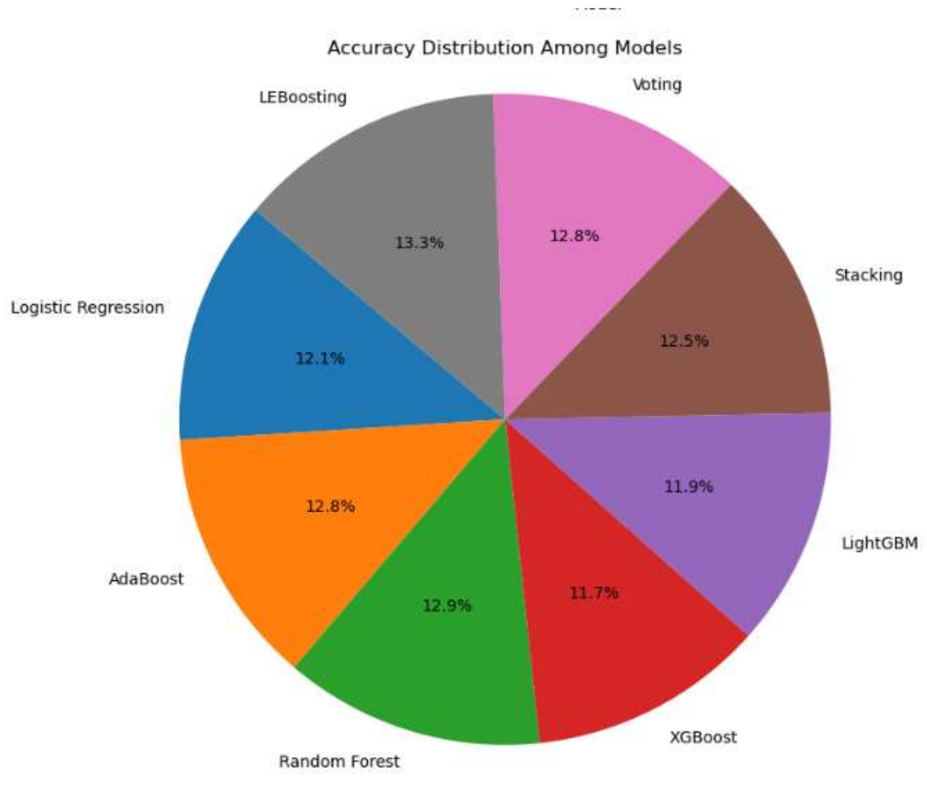

Figure 4: pie chart for used models in study

In their 2024 study, Narasimhan and Victor developed a novel approach for early detection of heart disease by combining Grey Wolf Optimization (GWO) with a stacked ensemble learning strategy. The GWO algorithm was employed as a metaheuristic tool to optimize the hyperparameters of the models, while the stacked ensemble integrated predictions from multiple classifiers to boost accuracy and enhance the model's reliability [16].

4.3 Handling Imbalanced Data

The problem of the imbalanced datasets is important, especially in heart failure prediction. Best First feature selections may lead to discover important influential features, and that class level imbalance using kind of class balancers model improve the prediction 8 [21]. The BOO-ST (Boosting, SMOTE and Tomek links) method was proposed to handle imbalanced class in HF datasets[27].

The problem of class imbalance is intrinsic to real-world datasets such as clinical datasets and introduces numerous challenges in the classifier training process. As a result, classifiers exhibit less then optimal accuracy, precision, recall, and great degree of misclassification[28]. Different class balancing strategies are tried out by the authors: Undersampling, Random oversampling, SMOTE, ADASYN, SVM-SMOTE, SMOTEEN, and SMOTETOMEK. The (SMOTEEN) balance approach generally outperformed all other balancing methods regardless the of classifier or clinical dataset[4].

To tackle this imbalance problem, techniques like the Smote-Xgboost algorithm have been proposed. Information gain based feature selection method extracts the main features from the dataset in order to avoid model over fitting, and the Smote-Enn method is used to treat the unbalanced data to obtain Sample data with positive and negative categories around the same[26]. The model's performance degrades when the amount of available data is limited or imbalanced. Subsequently, generative adversarial networks (GANs) can be employed to enlarge small/unbalanced datasets, and then models such as 1D-CNN and Bi-LSTM will be trained based on the augmented datasets for heart disease diagnosis [29].

To address class imbalance in datasets, various techniques such as proportional splitting, resampling (both over- and under-sampling), and ensemble-based methods like EasyEnsemble and w-EasyEnsemble are commonly applied. Following preprocessing, supervised learning models may be deployed, with Recursive Feature Elimination (RFE) used to identify key features and base-level models trained on the selected subsets. Several studies have highlighted the effectiveness of Gradient Boosting Decision Trees (GBDT) both as robust base classifiers and as optimal components in sampling strategies for constructing balanced datasets [30].

5. CHALLENGES AND FUTURE DIRECTIONS

5.1 Current Limitations

Although much advancement has been made, lots of limitations exist in ensemble machine learning models for HF prediction. Researches have proved that precision of heart disease predictions can be very high, but still with room for improvements[11]. The class imbalance in clinical datasets raises challenging problems when training classifiers; accuracy, precision, recall and misclassification rate are also problematic [28].

Thereby one has to map the image data into a numerical representation in order to use that in decision trees. This involves utilizing the output predictions of several independent Convolutional Neural Networks (CNNs) as feature inputs for training decision tree models. As an image representation, the CNN-based approaches are generic classification strategies that are applicable to any image dataset[34]. But, it makes the modeling a bit more complex.

Under the high workload working pressure, clinicians often get overwhelmed by a large amount of clinical signs produced in ICU, which could cause delay of treatment, care quality less than desired, or even misclinical decision [32]. The accuracy of the model reduces with very little or an imbalanced data collected that needs to be further handled[29].

Although people have applied classical machine learning on prediction and classification in cardiovascular diseases, most of them are not accurate enough and time-consuming, resulted from the limitation of high-quality dataset, imbalanced sample, inefficient preprocessing for data, and the selection of the criteria of the population. These lead to overfitting or bias towards them in prediction models if their predictive signals are not well-modeled[35].

5.2 Future Research Directions

From the literature, a few interesting directions for future work can be identified. Comparison of methods for predicting heart failure indicates that many ensemble techniques are superior to those previously proposed, so their development is likely to be of value[12]. Early diagnosis of heart disease can contribute to early prevention, early detection and early treatment [3], and the objective prediction will also play a positive role in curbing the death and disability caused by the disease.

The exploration of ensemble machine learning approaches, model like bagging, to improve the prediction performance of QSAR methods is a promising direction[22]. Novel machine learning methods could advance CDV's medicine by offering more precise and individual prognostic predictions for patients with heart failure [6].

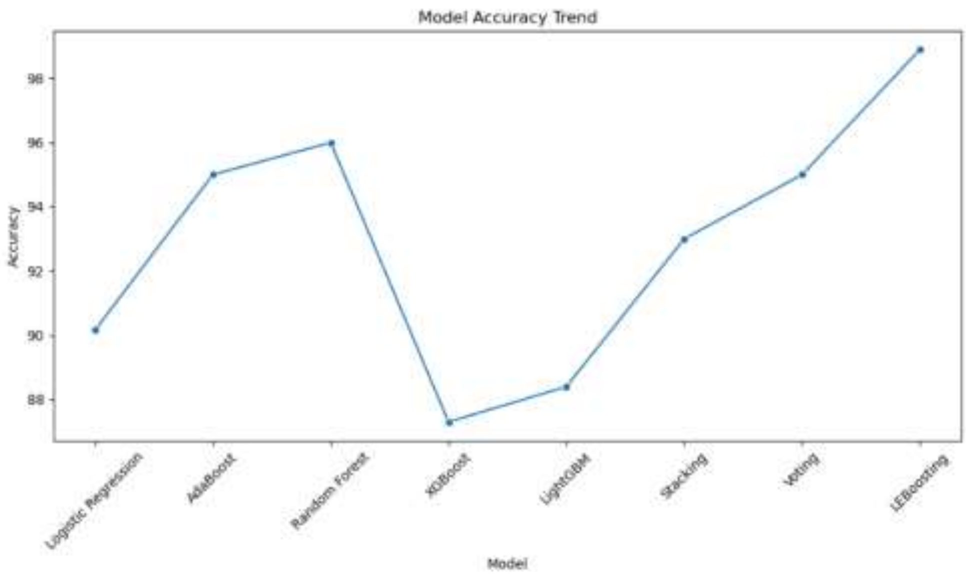

Figure 5: Model accuracy trends

Recently, new methods for fast detection and predicting heart diseases within few seconds using the-art-of-machine-learning, have been proven to significantly improve the predictive performance, and facilitating helpful support for medical personnel to quickly and more reliable evidence based decision [31]. New hybrid ML techniques such as BOO-ST and CBCEC may have a significant impact on saving life in heart failure patients and minimizing stress in the health care system[27].

The perceived efficiency of various ensemble methods on different canonical disease datasets varies. The knowledge of such variances could be helpful for researchers to choose alternative ensemble models for predictive disease analytics[9]. Using quantum inspired ML frameworks, medical data classification presents a potentially promising scenario even under the limitation of the existing resources. This course-work contains also significant contribution to the applications of quantum-based approaches, when signal is given data in medicine [36].

6. CONCLUSION

In this review, we have reviewed the existing literature on predicting HF using ensemble ML techniques. The results also indicate that ensemble methods achieve better performance than individual algorithms in heart failure prediction tasks, in which stacking and voting are the winners. These techniques have shown considerable potential for the detection of patients at risk of developing heart failure, which may allow for earlier intervention and improved outcomes.

There are a variety of ensemble methods that have been effectively used in practice, such as boosting-based algorithms such as AdaBoost, XGBoost, CatBoost, stacking (i.e., combining multiple models and feeding them into meta-learners), voting classifiers, which simply aggregate various model predictions, and random forest models that use ensembles of decision trees. There are also subjective comparisons as more novel algorithms, such as Least Error Boosting [20], BOO-ST [21], and CBCEC [12], are proposed for predicting heart failure and focus on the class imbalance and feature selection.

The performance of these models is highly competitive and the majority of the models achieved accuracies beyond 90%. Yet, challenges remain, e.g., data quality, imbalanced data, and the interpretability of complex ensemble models. In future studies, these limitations need to be mitigated, and the potential use of new techniques such as the quantum machine learning or transfer learning have to be investigated.

The clinical relevance of this study is substantial. Reliable predictive models may assist clinicians in early identification of patients at risk, optimizing resource allocation, and targeting interventions to reduce negative outcomes. As long as these models are refined and embedded in high-quality clinical decision support systems that are free of choice-supportive bias, they are set to revolutionize the management of heart failure that is overwhelming healthcare in many regions of the world.

REFERENCES

[1] Nissa, Najmu, Jamwal, Sanjay, and Neshat, Mehdi. 2024. "A Technical Comparative Heart Disease Prediction Framework Using Boosting Ensemble Techniques". De Computis. https://doi.org/10.3390/computation12010015

[2] Kedia, Shyam and Bhushan, Megha. 2022. "Prediction of Mortality from Heart Failure using Machine Learning". 2022 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET). https://doi.org/10.1109/ICEFEET51821.2022.9848348

[3] Asif, Daniyal, Bibi, M., Arif, M., and Mukheimer, A.. 2023. "Enhancing Heart Disease Prediction through Ensemble Learning Techniques with Hyperparameter Optimization". Algorithms. https://doi.org/10.3390/a16060308

[4] Chiu, Chih-Chou, Wu, Chung, Chien, Te-Nien, Kao, Ling-Jing, Li, Chengcheng, and Jiang, Hantao. 2022. "Applying an Improved Stacking Ensemble Model to Predict the Mortality of ICU Patients with Heart Failure". Journal of Clinical Medicine. https://doi.org/10.3390/jcm11216460

[5] Zhou, Dejia, Qiu, Hang, Wang, Liya, and Shen, Minghui. 2023. "Risk prediction of heart failure in patients with ischemic heart disease using network analytics and stacking ensemble learning". BMC Medical Informatics and Decision Making. https://doi.org/10.1186/s12911-023-02196-2

[6] Qadri, Azam Mehmood, Hashmi, Muhammad Shadab Alam, Raza, Ali, Zaidi, Syed Ali Jafar, and Rehman, A.. 2024. "Heart failure survival prediction using novel transfer learning based probabilistic features". PeerJ Computer Science. https://doi.org/10.7717/peerj-cs.1894

[7] Gutman, R., Aronson, D., Caspi, O., and Shalit, Uri. 2022. "What drives performance in machine learning models for predicting heart failure outcome?". European Heart Journal - Digital Health. https://doi.org/10.1093/ehjdh/ztac054

[8] Dritsas, Elias and Trigka, M.. 2023. "Efficient Data-Driven Machine Learning Models for Cardiovascular Diseases Risk Prediction". Italian National Conference on Sensors. https://doi.org/10.3390/s23031161

[9] Mahajan, P., Uddin, S., Hajati, F., and Moni, Mohammed Ali. 2023. "Ensemble Learning for Disease Prediction: A Review". Healthcare. https://doi.org/10.3390/healthcare11121808

[10] Laila, Umm-e, Mahboob, Khalid, Khan, Abdul Wahid, Khan, Faheem, and Taekeun, Whangbo. 2022. "An Ensemble Approach to Predict Early-Stage Diabetes Risk Using Machine Learning: An Empirical Study". Italian National Conference on Sensors. https://doi.org/10.3390/s22145247

[11] Ahmed, Sumaira, Shaikh, Salahuddin, Ikram, Farwa, Fayaz, Muhammad, Alwageed, H., Khan, Faheem, and Jaskani, Fawwad Hassan. 2022. "Prediction of Cardiovascular Disease on Self-Augmented Datasets of Heart Patients Using Multiple Machine Learning Models". Journal of Sensors. https://doi.org/10.1155/2022/3730303

[12] Chandrasekhar, Nadikatla and Peddakrishna, S.. 2023. "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization". Processes. https://doi.org/10.3390/pr11041210

[13] Mo, Yuhong, Zhang, Yuchen, Li, Hanzhe, Wang, Han, and Yan, Xu. 2024. "Prediction of heart failure patients based on multiple machine learning algorithms". Applied and Computational Engineering. https://doi.org/10.54254/2755-2721/75/20240498

[14] Sornsuwit, Ployphan, Jundahuadong, Phimkarnda, and Pongsakornrungsilp, Siwarit. 2022. "A New Efficiency Improvement of Ensemble Learning for Heart Failure Classification by Least Error Boosting". Emerging Science Journal. https://doi.org/10.28991/esj-2023-07-01-010

[15] Datta, Soumik, Hasan, S. M. Mahedy, Mitu, Mostarina, Taraque, Md Fakrul, Jannat, Nahrin, and Efat, Anwar Hossain. 2023. "Hyperparameter-Tuned Machine Learning Models for Complex Medical Datasets Classification". European Conference on Cognitive Ergonomics. https://doi.org/10.1109/ECCE57851.2023.10101525

[16] Narasimhan, Geetha and Victor, Akila. 2024. "Grey wolf optimized stacked ensemble machine learning based model for enhanced efficiency and reliability of predicting early heart disease". Automatika. https://doi.org/10.1080/00051144.2024.2317098

[17] Ambrews, Ashley Bryan, Moung, Ervin Gubin, Farzamnia, A., Yahya, Farashazillah, Omatu, S., and Angeline, Lorita. 2022. "Ensemble Based Machine Learning Model for Heart Disease Prediction". 2022 International Conference on Communications, Information, Electronic and Energy Systems (CIEES). https://doi.org/10.1109/CIEES55704.2022.9990665

[18] Munandar, Arif, Baihaqi, Wiga Maulana, and Nurhopipah, Ade. 2023. "A Soft Voting Ensemble Classifier to Improve Survival Rate Predictions of Cardiovascular Heart Failure Patients". Ilkom Jurnal Ilmiah. https://doi.org/10.33096/ilkom.v15i2.1632.344-352

[19] Coşkun, Cevdet and Kuncan, Fatma. 2022. "EVALUATION OF PERFORMANCE OF CLASSIFICATION ALGORITHMS IN PREDICTION OF HEART FAILURE DISEASE". Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi. https://doi.org/10.17780/ksujes.1144570

[20] Hassan, Ch. Anwar Ul, Iqbal, Jawaid, Irfan, Rizwana, Hussain, Saddam, Algarni, Abeer D., Bukhari, Syed Nisar Hussain, Alturki, Nazik, and Ullah, Syed Sajid. 2022. "Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers". Italian National Conference on Sensors. https://doi.org/10.3390/s22197227

[21] Yuliani, Yuri. 2022. "Algoritma Random Forest Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Seleksi Fitur Bestfirst". Infotek : Jurnal Informatika dan Teknologi. https://doi.org/10.29408/jit.v5i2.5896

[22] Enad, Huda Ghazi and Mohammed, M. A.. 2024. "Cloud computing-based framework for heart disease classification using quantum machine learning approach". Journal of Intelligent Systems. https://doi.org/10.1515/jisys-2023-0261

[23] Parikh, R., et al.. 2023. "Developing Clinical Risk Prediction Models for Worsening Heart Failure Events and Death by Left Ventricular Ejection Fraction". Journal of the American Heart Association : Cardiovascular and Cerebrovascular Disease. https://doi.org/10.1161/JAHA.122.029736

[24] Chen, Shengyue, Pan, Xinling, Mo, Jiahang, and Wang, Bin. 2023. "Establishment and validation of a prediction nomogram for heart failure risk in patients with acute myocardial infarction during hospitalization". BMC Cardiovascular Disorders. https://doi.org/10.1186/s12872-023-03665-2

[25] Xu, Liping, Cao, F., Wang, Lian, Liu, Weihua, Gao, M., Zhang, Li, Hong, Fuyuan, and Lin, Miao. 2024. "Machine learning model and nomogram to predict the risk of heart failure hospitalization in peritoneal dialysis patients". Renal Failure. https://doi.org/10.1080/0886022X.2024.2324071

[26] Yang, Jian and Guan, Jinhan. 2022. "A Heart Disease Prediction Model Based on Feature Optimization and Smote-Xgboost Algorithm". Information. https://doi.org/10.3390/info13100475

[27] Sutradhar, Ananda, et al.. 2023. "BOO-ST and CBCEC: two novel hybrid machine learning methods aim to reduce the mortality of heart failure patients". Scientific Reports. https://doi.org/10.1038/s41598-023-48486-7

[28] Kumar, Vinod, Lalotra, G. S., Sasikala, Ponnusamy, Rajput, D., Kaluri, Rajesh, Lakshmanna, Kuruva, Shorfuzzaman, Mohammad, Alsufyani, A., and Uddin, Mueen. 2022. "Addressing Binary Classification over Class Imbalanced Clinical Datasets Using Computationally Intelligent Techniques". Healthcare. https://doi.org/10.3390/healthcare10071293

[29] Sarra, Raniya R., Dinar, Ahmed M., Mohammed, M., Ghani, Mohd Khanapi Abd., and Albahar, M.. 2022. "A Robust Framework for Data Generative and Heart Disease Prediction Based on Efficient Deep Learning Models". Diagnostics. https://doi.org/10.3390/diagnostics12122899

[30] Liu, Ran, Wang, Miye, Zheng, T., Zhang, Rui, Li, Nan, Chen, Zhongxiu, Yan, Hongmei, and Shi, Qingke. 2022. "An artificial intelligence-based risk prediction model of myocardial infarction". BMC Bioinformatics. https://doi.org/10.1186/s12859-022-04761-4

[31] Pande, P., Khobragade, Prashant K., Ajani, Samir N., and Uplanchiwar, Vaibhav P.. 2024. "Early Detection and Prediction of Heart Disease with Machine Learning Techniques". 2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET). https://doi.org/10.1109/ICICET59348.2024.10616294

[32] Peng, S., et al.. 2022. "Interpretable machine learning for 28-day all-cause in-hospital mortality prediction in critically ill patients with heart failure combined with hypertension: A retrospective cohort study based on medical information mart for intensive care database-IV and eICU databases". Frontiers in Cardiovascular Medicine. https://doi.org/10.3389/fcvm.2022.994359

[33] Akkur, Erkan. 2023. "Prediction of Cardiovascular Disease Based on Voting Ensemble Model and SHAP Analysis". Sakarya University Journal of Computer and Information Sciences. https://doi.org/10.35377/saucis...1367326

[34] Khozeimeh, F., et al.. 2022. "RF-CNN-F: random forest with convolutional neural network features for coronary artery disease diagnosis based on cardiac magnetic resonance". Scientific Reports. https://doi.org/10.1038/s41598-022-15374-5

[35] Alghamdi, F., Almanaseer, Haitham, Jaradat, Ghaith M., Jaradat, Ashraf, Alsmadi, M., Jawarneh, Sana, Almurayh, Abdullah, Alqurni, J., and Alfagham, Hayat. 2024. "Multilayer Perceptron Neural Network with Arithmetic Optimization Algorithm-Based Feature Selection for Cardiovascular Disease Prediction". Machine Learning and Knowledge Extraction. https://doi.org/10.3390/make6020046

[36] Ozpolat, Zeynep and Karabatak, M.. 2023. "Performance Evaluation of Quantum-Based Machine Learning Algorithms for Cardiac Arrhythmia Classification". Diagnostics. https://doi.org/10.3390/diagnostics13061099.