ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR

CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.01 (2026)

GRADIENT BOOSTING-BASED HEART FAILURE PREDICTION FOR IMPROVED MEDICAL DECISION MAKING

[1,*]Sumbul Azeem, [2]Shazia Javed,[3] Farzad Zandi ,[4] Iftikhar Naseer, [5] Shamim Akhter

[1,2] Department of Mathematics, Lahore College for Women University (LCWU), Jail Road, Lahore, Pakistan.

[3] Department of Mathematics and Computer Science, Arak Branch, Islamic Azad University, Arak, Iran.

[4]Faculty of Computer Science and Information Technology, Superior University, Lahore, Pakistan.

[5]School of Information Management, Minhaj University, Lahore, Pakistan.

Email: [1] Sumbul.azeem@lcwu.edu.pk ,[2] shazia.javed@lcwu.edu.pk, [3] zandi8farzad@gmail.com, , [4] iftikharnaseer@gmail.com , [5] shamfgcollege@gmail.com

****Corresponding Author: Sumbul Azeem, Sumbul.azeem@lcwu.edu.pk***

Abstract:
Heart failure is a life-threatening cardiovascular disease that has a high mortality rate, where the heart is unable to pump blood effectively. Early and precise prediction is essential for timely intervention and better patient outcomes. Machine learning has demonstrated potential in discriminating high-risk patients using clinical data, yet most of the literature applies a single algorithm, which restricts predictive accuracy. The proposed study uses the state-of-the-art gradient boosting algorithms XGBoost and CatBoost to predict heart failure. The suggested methodology involves preprocessing that involves dealing with missing values, eliminating outliers, and normalization of features to guarantee the model with reliable input. Experimental evidence has shown that XGBoost is better than CatBoost with an accuracy of 77.78% and a lower false negative rate, showing better capability of identifying at-risk patients. Incorporating these models into the clinical decision support systems can lead to better early detection, timely treatment, and probably mortality associated with heart failure.

Keywords: *CatBoost, Classification, Gradient Boosting, Heart Disease, Machine Learning, Risk Prediction, XGBoost.*

1. Introduction:

Heart failure is a significant health problem in the world, which has been affecting millions of individuals across the globe and has been a leading cause of morbidity and mortality [1]. The number of people living with heart failure in 2021 was estimated to be more than 64 million and this number claims about 300,000 lives each year in the United States alone [2]. Heart failure is a condition that results when the heart cannot pump enough blood to supply the body with metabolic needs, which causes the body to build up fluid, dysfunction of organs, and a gradual deterioration of the heart. The problem with early diagnosis is that the first symptoms, e.g. fatigue or mild shortness of breath, are too common and may be easily missed and many patients are only diagnosed at a later stage with a worse prognosis [3].

The conventional techniques of diagnosing heart failure are based on the clinical examination, imaging tools such as echocardiography, and biomarkers such as the B-type natriuretic peptide

(BNP) and N-terminal proBNP (NT-proBNP) [4]. Although these methods have proven useful in the clinical setting, they are time-consuming, resource-intensive, and rely on the expertise of the physician that can slow down early diagnosis. The recent developments in artificial intelligence (AI) and machine learning (ML) have shown the possibility to improve the early diagnosis process and analyze complex clinical data to detect patterns that are not easily visible to human specialists [5]. In contrast to traditional rule-based systems, AI systems have the ability to learn latent associations between patient characteristics and outcomes, which can be used to predictive model timely intervention. Machine learning has already demonstrated positive outcomes in cardiovascular diagnostics, such as arrhythmia detection, risk prediction of myocardial infarction, and automated echocardiogram interpretation [6].

The prevalence and mortality of heart failure is high, and it is critically important to develop accurate, robust, and interpretable predictive models. XGBoost and CatBoost are examples of gradient boosting algorithms that have become strong classification algorithms because they are capable of nonlinear relationships, missing values, and feature interactions. The models have the potential to enhance the early detection, patient management and minimize adverse outcomes in clinical practice.

Major contribution of this paper are listed below:

- This paper will compare the XGBoost and CatBoost performance on heart failure prediction.
- The models facilitate early identification of high-risk patients, which helps in the timely interventions.
- The study brings out the clinical applicability of gradient boosting models in decision support.
- The research forms a basis of extension in the future, such as bigger datasets or mixed methods.

Section 2 of the article is based on a literature survey. Section 3 contains research methodology. Section 4 presents and discusses the results. Section 5 concludes the article and provide future directions.

2. Literature Review:

Following table 1, contains literature survey related to heart failure prediction.

Reference	Main Model	Metrics	Purpose
Zeng [7]	Decision Tree, Support Vector Machine, K nearest neighbors, XGBoost	Accuracy, Recall, F-1 Score.	Prediction of Heart Disease from Clinical Data
Wang [8]	Lasso-Logistic, XGBoost, Random Forest, K-Nearest Neighbor, and Support Vector Machine	AUC, Accuracy, Precision, F1 score, and Brier score	Prediction of in-hospital HFpEF risk
Hajishah[9]	Systematic Reviews and Meta-Analyses	Not specified	Assess ML models for Heart Failure outcomes
Kok [10]	Random Survival Forest (RSF) and	C-Index	Comparison of Machine Learning

	eXtreme Gradient Survival Boosting (XGBoost)		Models for incident Heart Failure risk
Mushtaq [11]	XGBoost + SHAP	F1-score, accuracy	Explainable prediction framework
Tanaka [12]	XGBoost interpretability	Not specified	Predict worsening Heart Failure
Mokhtari[13]	Graph Neural Network + explainability	Performance on echo data	Estimate Ejection Fraction for Heart Failure risk assessment

Table 1: Literature Survey

3. **Research Methodology:**

The rapid development of artificial intelligence has facilitated the medical diagnosis process by making it more accurate, efficient, and available. Machine learning algorithms are some of the most useful AI methods for identifying diseases early in the case of a patient by examining medical records [14]. Timely detection of heart failure risk is essential in the area of cardiovascular health to enhance patient outcomes. Gradient boosting algorithms and other machine learning models can be used to help physicians identify heart failure in its early stages with high accuracy. This paper presents a machine learning-based heart failure prediction system (**Figure 1.**) that uses a sophisticated gradient boosting algorithm, such as XGBoost and CatBoost, to produce reliable predictions.

The proposed model accepts the medical data of the patient as input and processes it initially by the preprocessing phase. In preprocessing, the missing values are addressed with the help of the corresponding algorithmic methods, and the outliers are identified and eliminated. This data is then normalized so that there is consistency across features. The second stage involves training the predictive models using the already processed data. The model was trained in terms of advanced gradient boosting algorithms, such as XGBoost and CatBoost, in order to learn more complex patterns in the data. During the last stage, the trained models are used to forecast the predictions using the pre-processed data of the patients. In case a patient is predicted to be in jeopardy of developing heart failure, a patient can be referred to a cardiologist to undergo additional tests and take early actions.

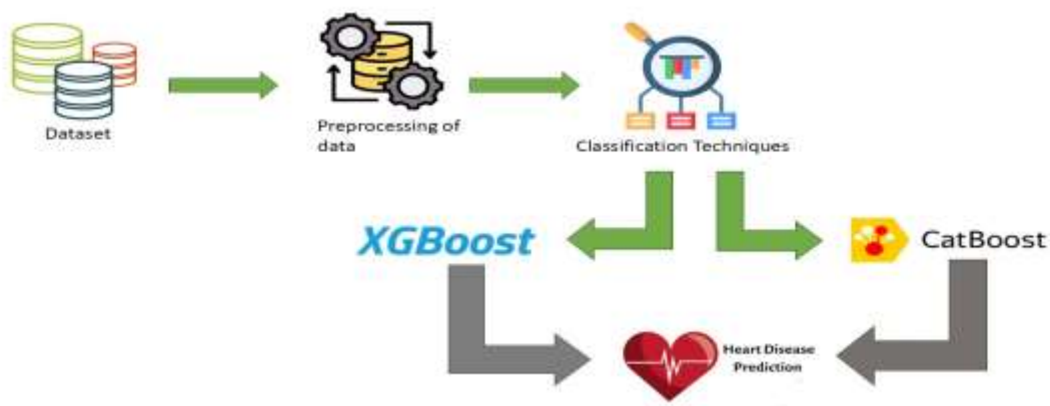


Figure 1: Proposed Methodology

3.1 DATSET:

The offered model was based on a heart failure prediction dataset obtained in Kaggle [15] that is free to access. In this data, there are 299 patients which include 13 independent variables, and these variables are age, ejection fraction, serum creatinine, and platelets and 1 dependent variable that shows the occurrence of death. The data is separated into two categories: 96 patients that had a death event (class 1) and 203 patients that did not (class 0). To train and predict the data, the suggested model had used different approaches to deal with missing values, identify and eliminate outliers, and to normalize the features. Histogram of numerical features are shown in Figure 2.

3.2 XGBoost

XGBoost short form for eXtreme Gradient Boosting is an innovative machine learning algorithm intended for efficacy, rapidity and high performance [16]. XGBoost constructs its models with decision trees as base learners and its construction (combination) by adding one after another in an attempt to improve overall performance. The tree is trained on the errors of the last tree and it is a process referred to as boosting. Parallel processing is also supported by the algorithm, which allows it to train on a large dataset. XGBoost, also, has a large amount of customization, and the model parameters can be fine-tuned by a user to a model that is optimally suited to a particular problem.

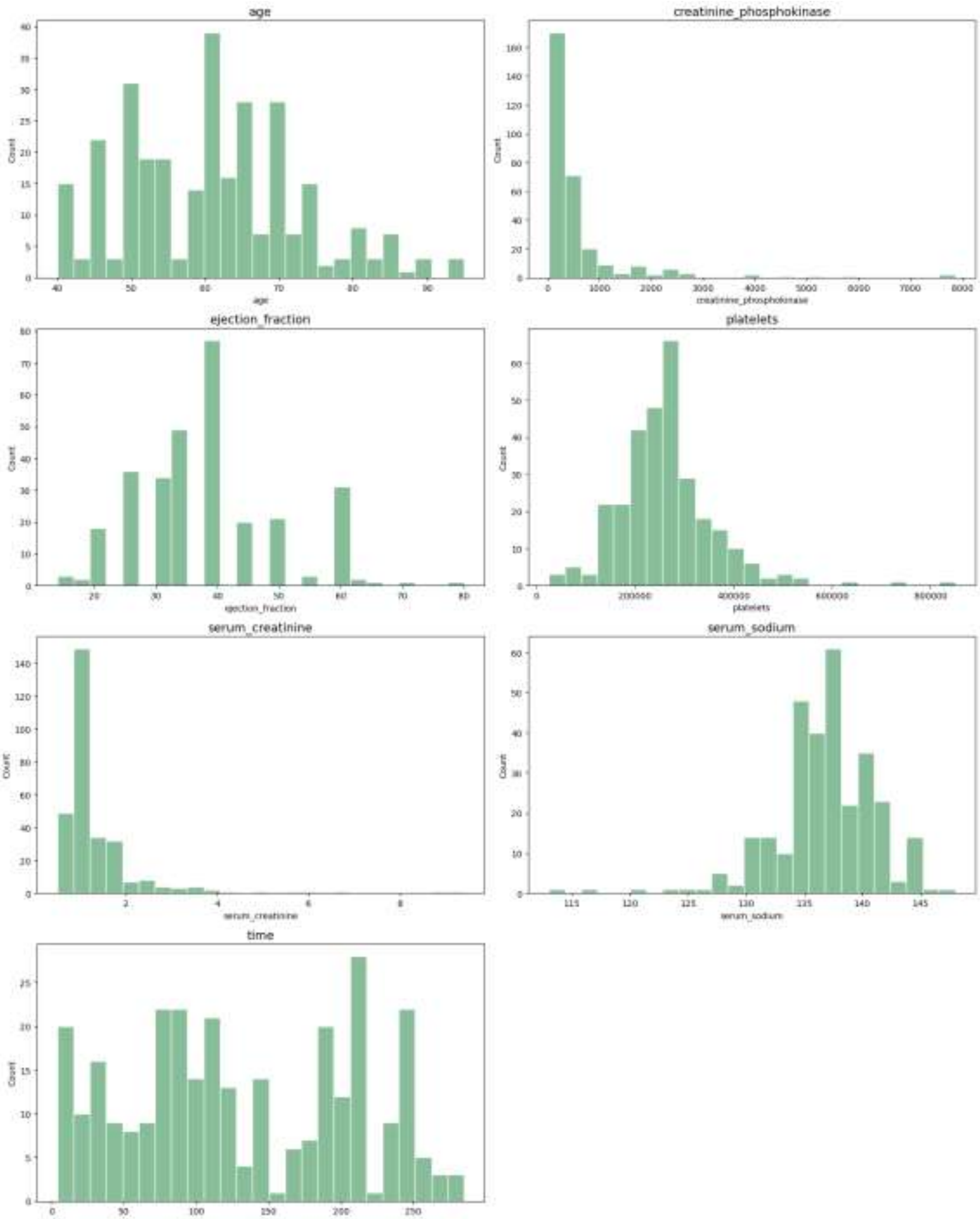


Figure 2: Histogram of Numerical Features

3.3 CatBoost

CatBoost (Categorical Boosting) is a gradient boosting algorithm, which is an ensemble of decision trees and is tailored to effectively work with the categorical features [17]. Compared to the traditional gradient boosting algorithms, which use a lot of preprocessing, CatBoost uses an ordered target encoding strategy, which minimizes target leakage and overfitting. The trees are built in sequence and each tree is trying to minimize the errors of the previous ensemble and this results in a better predictive performance.

3.4 Objective function:

The objective function of CatBoost and XGBoost contains two functions: a loss and regularization. The loss term measures the quality of the model on the training data and the regularization term ensures that the model is not over-fitted by balancing the complexity of the trees [18]. The overall loss function has the form:

$$obj(\phi) = \sum_{j}^{n} l(u_i \hat{u}_i) + \sum_{i=1}^{I} \psi(f_i) ---------- \mathbf{eq.(1)}$$

where

- $l(u_i \hat{u}_i)$is the loss function describing the difference between actual value u_i and predicted value $\hat{u}_i$.
- $\psi(f_i)$ is the regularization term which dejects overly intricate trees.

3.5 *Performance Metrics*

The data is split into 70:30 for training and testing. Random seed is set to 42. To check the authenticity and performance of the suggested model, different statistical parameters, including Classification accuracy, True Negative Rate (Specificity), True Positive Rate (Sensitivity), F-1 score, False Positive Rate (1-Speficity) and False Negative Rate (1- Sensitivity), are used. In the projected model attained, results, *p* represents true negative results, *q* represents false-positive results, *r* represents true positive results, and s represents false-negative results. They are described in equations (2) – (7).

$$\text{Accuracy} = \frac{\mathrm{p+r}}{\mathrm{p+q+r+s}} ---------- \mathbf{eq.(2)}$$

$$\text{Specificity} = \frac{\mathrm{p}}{\mathrm{p+q}} ---------- \mathbf{eq.(3)}$$

$$\text{Sensitivity} = \frac{\mathrm{r}}{\mathrm{r+s}} ---------- \mathbf{eq.(4)}$$

$$\text{F1-score} = \frac{\mathrm{2r}}{\mathrm{2r+q+s}} ---------- \mathbf{eq.(5)}$$

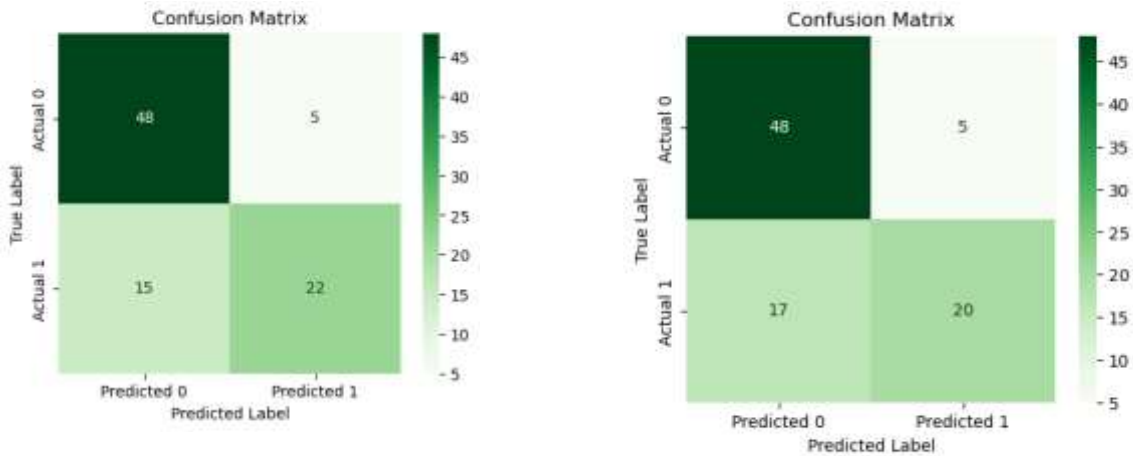
$$\text{False Positive Rate} = \frac{\mathrm{q}}{\mathrm{p+q}} ---------- \mathbf{eq.(6)}$$

$$\text{False Negative Rate} = \frac{s}{r+s} ---------- \mathbf{eq.(7)}$$

4. Simulation Results:

In this section, the obtained outputs and the values of the accuracy are discussed and shown. As a rule, the algorithm with the best accuracy is regarded as giving the best results. Figure 3 shows confusion matrices for both the algorithms.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR



XGBoost CatBoost

Figure 3: Confusion Matrix

The results of the accuracy of each of the evaluated algorithms are provided in Table 2.

Algorithm Metric	**XGBoost %**	**CatBoost %**
Accuracy	**77.78**	75.56
Sensitivity	**59.46**	54.05
Specificity	**90.57**	**90.57**
F1-score	**68.75**	64.52
False Positive Rate	**9.43**	**9.43**
False Negative Rate	**40.54**	45.95

Table 2: Simulation Results

The obtained experimental outcomes prove the predictive value of gradient boosting models in heart failure classification. Table I indicates that the XGBoost model was the best in terms of overall accuracy of 77.78 compared to the CatBoost model which had an accuracy of 75.56. This shows that XGBoost offers a little more dependable forecast on the assessed data. Regarding sensitivity, the capabilities of a model to identify correctly patients with heart failure, XGBoost demonstrated a higher true positive rate (59.46) than CatBoost (54.05%). It is a clinically valid improvement because an increase in sensitivity will decrease the chance of false negatives. The specificity of both models was the same, 90.57% which indicated that both models have good performance in terms of the ability to correctly identify non-heart-failure cases. The F1-score, which is a weighted mean of precision and recall, also supports the fact that XGBoost (68.75%) is better than CatBoost (64.52%). Also, the false positive rate in both models was low (9.43%), which means that the models were consistently reliable when it comes to negative prediction. Nevertheless, CatBoost had a higher false negative rate, which means that it was more likely to classify the positive cases as negative. In general, both gradient boosting methods proved to be effective in terms of classification, but XGBoost proved to be a more balanced and robust at the same time, which is why it is more applicable to heart failure predictions in this paper. Table 3 shows comparative analysis of proposed technique with some of those already in literature.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.04 No.01 (2026)

Model	Model	Accuracy (%)
Proposed Method	XGBoost CatBoost	77.78 75.56
Qadri [19]	Support Vector Machine K Nearest Neighbors	72 58

Table 3: Comparative Analysis

5. Conclusion and Future Directions:

We compared the results of two gradient boosting algorithms, XGBoost and CatBoost, in predicting heart failure given patient clinical data in this study. The findings have shown that both models can be successfully used to classify patients, although XGBoost has more overall accuracy, sensitivity, and F1-score than CatBoost. The two models were also highly specific and low in false positive rates meaning that they are reliable in true detection of non-heart-failure cases. These results demonstrate that gradient boosting algorithms can be useful in heart failure early detection and help clinicians intervene in time and achieve better patient outcomes. There are a number of directions that can be pursued in future research. The generalizability of the models might be enhanced by incorporating bigger and more diverse datasets. The combination of gradient boosting with other machine learning or deep learning techniques can also be used as a hybrid approach that can improve predictive performance. Moreover, it may be better to incorporate explainable AI methods in order to enhance the interpretability of model predictions that are useful in clinical decision-making. Lastly, practical use of these models and constant monitoring of their results in a clinical environment would be necessary to prove their usefulness and influence on patient care.

References:

[1] Savarese, Gianluigi, and Lars H. Lund. "Global public health burden of heart failure." *Cardiac failure review* 3.1 (2017): 7.

[2] Virani SS, et al. “Heart Disease and Stroke Statistics-2021 Update: A Report From the American Heart Association”. Circulation. 2021 Feb 23;143(8):e254-e743. doi: 10.1161/CIR.0000000000000950. Epub 2021 Jan 27. PMID: 33501848.

[3] Meunier-McVey, Natasha. "2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure." *CARDIOLOGY* (2021): 23.

[4] Ponikowski, Piotr et al. “Heart failure: preventing disease and death worldwide.” *ESC heart failure* vol. 1,1 (2014): 4-25. doi:10.1002/ehf2.12005.

[5] Johnson, Kipp W., et al. "Artificial intelligence in cardiology." *Journal of the American College of Cardiology* 71.23 (2018): 2668-2679.

[6] Attia, Zachi I., et al. "Screening for cardiac contractile dysfunction using an artificial intelligence–enabled electrocardiogram." *Nature medicine* 25.1 (2019): 70-74.

[7] Zeng, Ming. "The Prediction of Heart Failure based on Four Machine Learning Algorithms." *Highlights in Science, Engineering and Technology* 39 (2023): 1377-1382.

[8] Wang, Jing-Xian et al. “Machine learning algorithms to predict heart failure with preserved ejection fraction among patients with premature myocardial infarction.” *Frontiers in cardiovascular medicine* vol. 12 1571185. 7 May. 2025, doi:10.3389/fcvm.2025.1571185.
[9] Hajishah, Hamed, et al. "Evaluation of machine learning methods for prediction of heart failure mortality and readmission: meta-analysis." *BMC Cardiovascular Disorders* 25.1 (2025): 264.
[10] Kok, Thomas F., et al. "High-dimensional machine learning models for prediction of heart failure in more than 400 000 men and women from the UK Biobank." *European Heart Journal-Digital Health* 6.6 (2025): 1234-1245.
[11] Mushtaq, Mehtab, et al. "Causal and Explainable Machine Learning Framework for Heart Disease Prediction using XGBoost and SHAP." *Journal of Computing & Biomedical Informatics* 10.01 (2025).
[12] Tanaka, Munekazu, et al. "Development of interpretable machine learning models to predict in-hospital prognosis of acute heart failure patients." *ESC Heart Failure* 11.5 (2024): 2798-2812.
[13] Mokhtari, Masoud, et al. "Echognn: Explainable ejection fraction estimation with graph neural networks." *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer Nature Switzerland, 2022.

[14] Azeem, Sumbul, Shazia Javed, and Simran Azeem. "Hybrid PSO-HHO Based Machine Learning Optimization For Lung Cancer Diagnosis." *2025 International Conference on Emerging Technologies in Electronics, Computing, and Communication (ICETECC)*. IEEE, 2025.
[15] Andrew Mvd, “Heart Failure Clinical Data,” Kaggle, 2020. Available: https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data/data
[16] Bentéjac, Candice, Anna Csörgő, and Gonzalo Martínez-Muñoz. "A comparative analysis of gradient boosting algorithms." *Artificial Intelligence Review* 54.3 (2021): 1937-1967.
[17] Murtiningsih, Dewi Ayu, Bety Wulan Sari, and Ika Nur Fajri. "Comparison of Light Gradient Boosting Machine, eXtreme Gradient Boosting, and CatBoost with Balancing and Hyperparameter Tuning for Hypertension Risk Prediction on Clinical Dataset." *Journal of Applied Informatics and Computing* 9.5 (2025): 2753-2763.
[18] Aish, Muhammad Abdullah, et al. "Improving stroke prediction accuracy through machine learning and synthetic minority Over-sampling." *Journal of Computing & Biomedical Informatics* 7.02 (2024).
[19] Qadri, Azam Mehmood et al. “Heart failure survival prediction using novel transfer learning based probabilistic features.” *PeerJ. Computer science* vol. 10 e1894. 12 Mar. 2024, doi:10.7717/peerj-cs.1894.