Federated Learning-Based Privacy-Preserving Hybrid Model for Heart Disease Prediction Using SVM and XGBoost

[1]Randhir Kumar, [2]Tarunendra Singh
[1]M. Tech Scholar, Department: Computer science engineering, Millennium Institute of technology
[2]Assistant professor, Department: Computer science engineering, Millennium Institute of technology
[1]randhir.surya1@gmail.com, [2]trs.singh89@gmail.com

* Corresponding Author: Randhir Kumar

Abstract:

With heart disease being the prime cause of death worldwide, there is an ever-pressing need for the establishment of worthwhile and foremost prediction methods. However, limitations arise due to privacy issues concerning sensitive patient data, which work against any centralized machine learning framework. Thus, the current paper sets forth a holistic privacy-preserving framework for heart disease prediction by Federated Learning (FL) integrating hybrid SVM (support vector machine) with an XGBoost model. The architecture allows various medical institutions (called Alice and Bob) to train their own SVMs locally on private datasets without the exchange of raw data. Instead, these local models share the parameter values of their models with a central aggregator, which in turn uses XGBoost to devise a global meta-learner capable of identifying both linear and non-linear patterns across distributed datasets. Key preprocessing phases include imputation of missing values through means, label encoding, min-max normalizing, and up-sampling for class balancing. The model is judged against reliable metrics like Accuracy, Precision, Recall, F1-score, and ROC-AUC, with k-fold cross-validation assuring robustness. The hybrid model can ensure greater generalizability and can withstand imbalanced-class situations much better than traditional methods. Association Rule Mining is then added to offer decision rules for clinical explainability. The method enables training SVM models locally with Alice and Bob both achieving high accuracies of 98.2% and 98.7%, with F1-scores of 0.904 and 0.905, respectively. However, both models shared similar recall values, approximately 0.89, which suggests false negatives, a major deterrent in medical diagnosis. Global classifier, the other hand, showed better performance: with an overall accuracy of 98.5%, precision 0.958, recall 0.936, and F1-score 0.947. This research is indicative of a practical approach toward decentralised, secured, and interpretable predictive analytics in healthcare. It demonstrates that federated frameworks with a high level of diagnostic accuracy can also claim privacy and data protection (e.g., HIPAA, GDPR) and thereby can present a practical, ethical approach toward analysing real-world medical data.

Keywords: *Heart Disease Prediction, Federated Learning, Privacy-Preserving AI, Support Vector Machine, XGBoost, Medical Data Security*

I. INTRODUCTION

Cardiovascular diseases still remain among the major causes of death, with millions dying each year. The earlier the diagnosis and intervention, the better the prospects for patients and fewer financial burdens on healthcare systems [1]. Thus, with the digitization of medical records and health data proliferating between institutions, there appears to be much untapped potential for the development of enhanced computational models for heart disease prediction. However, this data is often distributed across different hospitals and research centers, and barriers exist regarding data sharing, privacy, and interoperability [2]. In regular centralized machine learning approaches, raw data must be transmitted; but this is impossible because of regulatory and ethical issues. Federated learning and privacy-preserving data mining approaches appear to be potential solution avenues to overcome these obstacles [3]. This work is aimed at the development of an efficient heart disease prediction model from distributed media. Fig.1shows Heart Disease Prediction

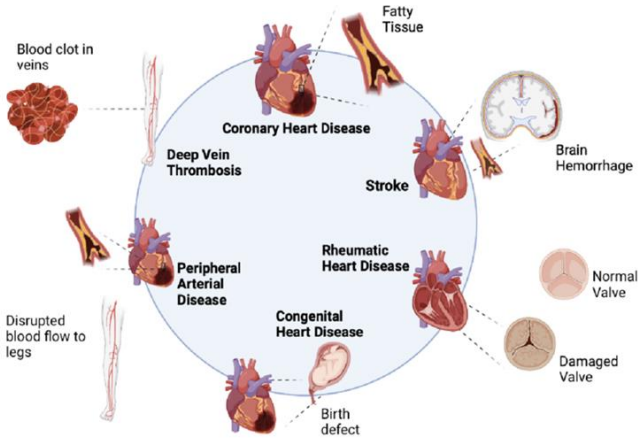

Fig.1: Heart Disease Prediction [3]

Cardiovascular illnesses rank first in terms of causing death, claiming nearly 17.9 million deaths yearly, which constitutes about 32% of all deaths worldwide, according to WHO. Heart diseases, including coronary artery disease [4], arrhythmias, and heart failure, are the most common among CVDs and are largely responsible for the socio-economic burden. Upward trends of heart-related ailments have been attributed to sedentary lifestyles, unhealthy food habits, obesity, diabetes, and smoking, especially in low- and middle-income countries [5]. On top of that comes the silent onset of early symptoms in most cases, making diagnosis at a very early stage almost impossible until symptoms are severe. Hence, the need for mechanisms that will help in cantilevering the disease early and an accurate prediction of the disease. Combining digital technologies with health—which include Electronic Health Records (EHR)—has become an ultimate answer to the realization of data-driven healthcare [6]. Machine learning has further entered this scene to train-medical data with high dimensions and uncover latent patterns in either clinical or ECG features or demographics using algorithms like Support Vector Machines (SVM), Decision Trees, Neural Transfer, and XGBoost. Thereby stratifying risks of patients, enabling preventive care, and facilitating personalised treatment choices when the choice is there on. But the course of success for machine learning in healthcare depends largely on access to large and diverse datasets, which are all too often split apart across multiple institutions [7].

Whilst the advantages are theoretically there, challenges abound for the centralized learning frameworks in medicine, mainly because patient data is so sensitive and heavily protected by privacy laws. Centralizing datasets from several different institutions increases the chances for data breach and identity theft and could raise even higher legal and ethical concerns [8]. The stigma around confidentiality often makes it very difficult for either patient groups or providers to proceed with sharing data, thereby posing the greatest challenge to assembling comprehensive training data sets. Apart from these, strict regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the U.S. and the General Data Protection Regulation (GDPR) [9] in the EU lay down stringent restrictions on how medical data can be shared and processes, particularly across borders. These hurdles therefore, impede the linking of data from multiple sources to aircraft training for accurate, and generalized ML models. On top of this, the technical inconveniences stare back: centralized solutions for model training often suffer from issues of scalability, latencies, and heavy infrastructure costs. Furthermore, the difference present between patient data structure and quality among various institutions stigmatizes integration even more [10]. In contrast, the traditional setting of data accumulation within a central repository for training has been somewhat identified as a risky and rather inefficient alternative in the case of healthcare settings.

With these issues, FL has developed into one such transformative stage, making it ideally suited to the healthcare industry. FL allows various medical institutions to train machine learning models collaboratively without having to share the dirty data. Instead, the institutions (clients) train locally on the private datasets and share only the learned parameters, such as weights or gradients, with the central server for aggregation. This decentralized setup complies with data protection laws and serves to minimize the risk associated with privacy issues [11] the aggregated global model is then sent back to the institutions for further local training, fostering an iterative and privacy-compliant learning process. FL really shines when it comes to settings such as heart disease prediction, where data lies in different healthcare centers, and patients differ based on demographics and clinical profiles. By enabling secure, distributed training, FL ensures model generalizability and performance while remaining compliant with HIPAA and GDPR. Therefore, FL is not only a technical innovation but also a practical and scalable ethical solution to developing robust predictive models in modern healthcare landscapes.

With FL securing distributed training, generalizability, and the performance of AI models, it guarantees the health information privacy act (HIPAA) and the General Data Protection Regulation (GDPR). In fact, this keeps FL from being seen merely as a technical innovation and situates it as a real-world, scalable, and morally appropriate solution to build powerful predictive models in modern healthcare ecosystems. In this light, the present review provides a comprehensive

discussion on recent machine learning-based approaches for predicting heart diseases, analyzing their methodology, datasets, performance criteria, and key limitations-laying the foundation for the incorporation of privacy-preserving technologies like FL into clinical use [12].

II. LITERATURE REVIEW

Chintan M. Bhatt et al. [1] (2023) performed K-modes clustering with Huang initialization and used four algorithms, namely Random Forest, Decision Tree, MLP, and XGBoost, on a Kaggle dataset consisting of 70,000 instances. Results of GridSearchCV on hyperparameter tuning showed MLP to be better than its counterparts, with classification accuracy reported at 87.28% and AUC scores reaching 0.95. Nevertheless, issues of generalizability arise owing to the size and nature of the dataset.

ALLE HARSHA VARDHAN et al. [2] (2023) proposed a hybrid ensemble classifier which combined strong and weak learners, utilizing a large training and validation dataset to improve prediction accuracy for heart conditions. The ensemble outperformed single classifiers such as Random Forest, Decision Tree, SVM, Naive Bayes, and Logistic Regression. However, the study does not mention any limitations of the model or its generalizability to datasets that are diverse.

Zeinab Noroozi et al. [3] (2023) have located and examined a set of sixteen feature selection methods and seven ML algorithms using the Cleveland Heart Disease dataset. The performance of J48 improved greatly through feature selection, but the accuracy for MLP and Random Forest deteriorated. The best accuracy observed was 85.5%, attained by SVM-CFS, Information Gain, and the like. This leads to a number of issues, especially because of the small size of the dataset restricting use of the models for actual clinical applications.

Nadikatla Chandrasekhar et al. [4] (2023) implemented six algorithms: Random Forest, KNN, Logistic Regression, Naïve Bayes, Gradient Boosting, and AdaBoost on Cleveland and IEEE Dataport datasets. Their ensemble-based approach using soft voting yielded better performance with an accuracy of 93.44% on Cleveland and 95% on IEEE Dataport. This, however, puts into question the applicability in real-life clinical settings due to dependence on curated datasets.

Qadri et al. [5] (2023) proposed a new method of feature engineering, Principal Component Heart Failure (PCHF), and tested it under nine machine learning algorithms. The Decision Tree model, in particular, managed to give 100% accuracy, showing its real potential. This method, however, may suffer from overfitting as it was tested on a dataset that is rather small or too specific.

Biswas et al. [6] (2023) attempted to perform feature selection by Chi-Square, ANOVA, and Mutual Information methods, using six classifiers. Random Forest combined with the mutual information subset SF3 gave the best result of 94.51% accuracy and AURC of 94.95. However, this also limited the generalization of the model across other diverse populations because of its dependency on a particular healthcare dataset.

K. Arumugam et al. [7] (2023) worked on diabetic-specialized heart disease prediction employing Decision Tree, Naive Bayes, and SVM classifications, wherein the Decision Tree performed slightly better than others. However, the study stands constrained due to the limited availability of complete datasets specific to diabetic patients.

Ahmed A. H. Alkurdi et al. [8] (2023) built the entire preprocessing pipeline for normalization, SMOTE, and feature selection with the UCI Heart Disease dataset. They evaluated Decision Trees, Random Forest, SVM, and k-NN classifiers, all highly capable of metrics such as accuracy and ROC AUC. The biggest disadvantage is that it has become overly dependent on SMOTE and thus, might be an introduction for synthetic bias.

Mr. J. A. Jevin et al. [9] (2023) proposed a distributed association rule mining framework utilizing intelligent agents across different medical data sites under privacy constraints. Under stringent privacy considerations, the framework enabled the efficient discovery of global rules with very low communication. The drawback, however, is that coordination of agents becomes rather difficult in heterogeneous and dynamic environments.

K-modes clustering and machine-learning models like Random Forest, Decision Tree, MLP, and XGBoost were applied by Mukesh Kumar Saini et al. [10] (2023) on a Kaggle Dataset of 70,000 samples to GRIDSEARCHCV for tuning of parameters. MLP reported accuracy of 87.28% with good AUC scores though dependence on a single dataset undermines the utility of the model in the real world.

M. H. Fadly et al. [15] (2023) applied SVM, AdaBoost, and hybrid SVM-AdaBoost models on the UCI Cardiac Disease dataset based on the CRISP-DM methodology. The hybrid technique obtained 90% accuracy, which was better than what

SVM and AdaBoost achieved individually. Nevertheless, there was not any external validation, so the method cannot be generalized to broader clinical environments.

S. Yuda Prasetyo et al. [16] (2023) analyzed SVM, Naive Bayes, Decision Tree, and Random Forest algorithms on the Heart Failure Prediction dataset. Random Forest (91.85%) and SVM (90.76%) showed promising results, thereby justifying their use for heart disease risk prediction. Nonetheless, the study requires further tuning and validation on a larger dataset.

H. V. R. Bindela et al. [17] (2023) applied SVM with an RBF kernel and K-means clustering on the UCI Cardiac Disease dataset. SVM scored 91.85% accuracy, while K-means was able to segregate some subgroups with an accuracy of 84%. The primary drawback is the manual setting of the number of clusters, which decreases consistency and scalability.

[Six ML models, namely Logistic Regression, SVM, Decision Tree, Bagging, XGBoost, and LightGBM, were assessed for the prediction of myocardial disease by J. Miah et al. 18] (2023). XGBoost scored first with 92.72% accuracy. The lack of external data testing curbed the robustness of the model.

Anudeepa Gon et al. [19] (2023) examined whether Neural Networks, Logistic Regression, SVM, Random Forest, Naive Bayes, AdaBoost, and XGBoost yield improvements when applied to clinical and demographic features. Hence, different versions of the system could attain great accuracy, thus promoting early detection through feature importance insights. On the other hand, applicability to new populations now depends on the quality and scope of the training data.

V. R. Burugadda et al. [20] (2023) worked with Logistic Regression, Decision Tree, Random Forest, SVM, and ANN methods and predicted heart failure readmissions using EHR data. The models helped identify the patients at a high risk of readmissions to better plan their interventions. Limitations include some lack of interpretability and the underrepresenting of some socioeconomic variables.

In 2024, S. NagaMallik Raj et al. [21] designed a web application thatIntegrated XGB-Classifier and gradient boosting are applied on UCI Heart Disease dataset. With an accuracy of 85% and 93%, the system offers reliable risk predictions, allowing the users to evaluate the risk adequately. However, the existing prediction model does not consider the time-based features and may be overfitting; therefore, restricting its wider applicability.

In 2024, Sarah A. Alzakari et al. [22] integrated an IoT-system with XGBoost and Bi-LSTM models for remote monitoring of cardiac diseases with the real-time and electronic clinical data. This framework produces 99.4% accurate prediction with the best temporal forecast, but privacy issues and challenges in deploying it on a large scale come up as major obstacles.

J. Shanker Mishra et al. [23] (2024) took advantage of XGBoost, Bi-LSTM, and ResNet for cardiac datasets and MRI imaging to achieve an enhanced diagnostic accuracy of up to 99.4%. The Deep learning inclusion in the system increased the capability for enhancing feature representation while still requiring solutions for model interpretability and annotated data.

H. F. El-Sofany et al. (2024) [24] theoretically used feature selection (Chi-square, ANOVA, Mutual Information), combined with ten ML models, including XGBoost and the SVM, on the UCI dataset. With the SF-2, XGBoost attained an accuracy of 97.57% and an AUC of 98% according to SHAP interpretation. Still, clinical validation is lacking, and this brings the synthetic data bias into question.

Class imbalance was tackled by Adedayo Ogunpola et al. [25] (2024) through optimizations of XGBoost, CNN, Random Forest, and various other classifiers on the UCI dataset. XGBoost topped the leaderboards with 98.50% accuracy and 98.71% F1 score. While the results are quite good, one wonders whether the said results will generalize induced because the tuning was done in a specific dataset.

Table 1: Based on Machine Learning Techniques

Ref (Author, Year)	Technique Used	Dataset Used	Key Findings	Results	Limitations
[1] Chintan M. Bhatt et al., 2023	K-modes clustering with Huang initialization + RF, DT, MLP, XGB	Kaggle dataset (70,000 instances)	GridSearchCV tuning improves classification. MLP outperforms others.	MLP: 87.28% accuracy; AUC up to 0.95	Limited generalizability due to dataset size and composition

[2] ALLE HARSHA VARDHAN et al., 2023	Hybrid Ensemble Classifier integrating weak and strong learners	Large training and validation datasets	Ensemble model outperforms individual models in predicting heart conditions	Ensemble > RF, DT, SVM, NB, LR in accuracy	Not specified; possibly generalizability not discussed
[3] Zeinab Noroozi et al., 2023	16 Feature Selection Methods + 7 ML algorithms	Cleveland Heart Disease Dataset	Feature selection boosts J48 performance but reduces MLP and RF	Accuracy up to 85.5% with SVM-CFS, Info Gain	Small dataset, limited real-world applicability
[4] Nadikatla Chandrasekhar et al., 2023	RF, KNN, LR, NB, GB, AdaBoost + Soft Voting Ensemble	Cleveland & IEEE Dataport datasets	Ensemble outperforms individual models	Soft Voting: 93.44% (Cleveland), 95% (IEEE Dataport)	Dependency on curated datasets
[5] A. M. Qadri et al., 2023	PCHF feature engineering + 9 ML algorithms	Health data (dataset unspecified)	DT achieves perfect classification; PCHF improves detection	DT: 100% accuracy	Overfitting due to small or specific dataset
[6] Niloy Biswas et al., 2023	Chi-Square, ANOVA, Mutual Info + 6 ML classifiers	Not specified (healthcare dataset)	RF with mutual info features (SF3) performs best	RF: 94.51% accuracy, 94.95 AURC	Dataset dependency limits broad generalization
[7] K. Arumugam et al., 2023	Decision Tree, Naive Bayes, SVM	Diabetes-specific heart disease dataset	DT performs best in diabetic heart disease prediction	DT > SVM, NB (accuracy not specified)	Limited diabetic-specific data
[8] Ahmed A. H. Alkurdi et al., 2023	DT, RF, SVM, k-NN + Preprocessing (SMOTE, Normalization, Feature Selection)	UCI Heart Disease Dataset	Robust preprocessing pipeline enhances model performance	High scores across all metrics (Accuracy, Precision, ROC AUC)	Overuse of SMOTE may cause synthetic bias
[9] Mr. J. A. Jevin et al., 2023	Distributed Association Rule Mining using Multi-Agent System	Distributed medical data (privacy constraints)	Localized computation enables privacy-preserving rule mining	Efficient rule discovery with minimal communication	Complexity in agent coordination in dynamic networks
[10] Mukesh Kumar Saini et al., 2023	K-modes clustering + RF, DT, MLP, XGB + GridSearchCV	Kaggle (70,000 instances)	MLP achieves highest accuracy; strong AUC values for all	MLP: 87.28% accuracy, AUC up to 0.95	Single dataset limits cross-scenario applicability
[15] M. H. Fadly et al. (2023)	SVM, AdaBoost, Hybrid (SVM-AdaBoost)	UCI Cardiac Disease Dataset	Hybrid model offers best performance using CRISP-DM methodology	Hybrid: **90%**, SVM & AdaBoost: **86.67%**	No external validation; limits generalizability
[16] S. Yuda Prasetyo et al. (2023)	SVM, Naive Bayes, Decision Tree, Random Forest	Heart Failure Prediction Dataset	RF and SVM showed strong accuracy for heart disease risk prediction	RF: **91.85%**, SVM: **90.76%**	Needs further tuning and testing on larger datasets

[17] H. V. R. Bindela et al. (2023)	SVM (RBF), K-means Clustering	UCI Cardiac Disease Dataset	High SVM accuracy; K-means finds hidden subgroups	SVM: **91.85%**, K-means: **84%**	Manual cluster selection limits consistency
[18] J. Miah et al. (2023)	LR, SVM, DT, Bagging, XGBoost, LightGBM	Not UCI Cardiac Disease Dataset	XGBoost outperformed others in myocardial illness prediction	XGBoost: **92.72%**, LightGBM: **90.60%**	No external validation reduces robustness
[19] Anudeepa Gon et al. (2023)	Neural Networks, LR, SVM, RF, NB, AdaBoost, XGBoost	Clinical & Demographic Data	High accuracy; feature importance helps in early detection	High accuracy (not quantified)	Real-world applicability depends on dataset quality
[20] V. R. Burugadda et al. (2023)	LR, DT, RF, SVM, ANN	Electronic Health Records (EHR)	ML models help identify high-risk heart failure readmission patients	Evaluated via accuracy, precision, recall, F1	Gaps in interpretability and fairness due to unbalanced features
[21] S. NagaMallik Raj et al. (2024)	XGB-Classifier, Gradient Boosting	UCI Heart Disease Dataset	Web app enables early diagnosis and risk prediction	XGB: **85%**, GB: **93%**	Excludes time-based feature; possible overfitting
[22] Sarah A. Alzakari et al. (2024)	IoT + XGBoost + Bi-LSTM	ECD + Real-time Data	Remote monitoring with Bi-LSTM yields excellent temporal prediction	**Accuracy: 99.4%**	Privacy and IoT deployment challenges
[23] J. Shanker Mishra et al. (2024)	XGBoost, Bi-LSTM, ResNet	Cardiac Data + MRI Images	Combines imaging and structured data; deep learning boosts accuracy	Accuracy: **up to 99.4%**	Needs annotated data; interpretability concerns
[24] H. F. El-Sofany et al. (2024)	FS (Chi2, ANOVA, MI) + 10 ML Models incl. XGBoost, SVM, RF	UCI Cardiac Disease Dataset	XGBoost with SF-2 subset gave top accuracy; SHAP for explainability	Accuracy: **97.57%**, AUC: **98%**	Lacks clinical validation; synthetic data may bias results
[25] Adedayo Ogunpola et al. (2024)	XGBoost, CNN, RF, + 4 others	UCI Cardiac Disease Dataset	Tackles class imbalance; XGBoost achieved best overall metrics	Accuracy: **98.50%**, F1: **98.71%**	Limited generalizability beyond tuned dataset

III. RESEARCH OBJECTIVES

- To develop a privacy-preserving federated learning framework for heart disease prediction using distributed medical datasets.
- To train local Support Vector Machine (SVM) models at each institution without sharing raw patient data. To address class imbalance through up-sampling and advanced modeling techniques.
- To aggregate local model parameters to build a robust global model using XGBoost.
- To validate model generalizability and prevent overfitting using k-fold cross-validation.

IV. RESEARCH METHODOLOGY

A. Distributed Association Rule Mining algorithm for Predicting heart diseases

The DARM algorithm extracts hidden/unseen patterns from distributed medical databases with privacy preservation created through encrypted summary statistics instead of sharing patient-level data.These outcomes are studied through interestingness measures such as support, confidence, and lift- not prediction metrics such as precision or ROC-AUC.The generalizability of the model is tested by validating association rules in multiple data silos, with different distributions.Even though this approach retains privacy, it does not retain the flexibility and discriminating potential of supervised machine learning for complex, non-linear prediction tasks. Privacy Preservation with the proposed methodology, privacy preservation is a primary concern given the sensitive nature of medical data. To keep patient data private, federated learning architecture is used to combine both SVM and XGBoost models so that the data never gets shared across institutions.In this methodology, the local SVM model is trained at each institution on its own data, therefore learning the model without any raw data leaving the institution to some central place.

B. SVM model For Predicting Heart Diseases

This federated framework is designed for two medical institutes so that they can train models for heart disease prediction by sharing the model parameters only, such as support vectors and hyperplane coefficients, instead of actual patient data in compliance with the GDPR/HIPAA data privacy requirements, and also privacy is maintained by cryptographic means. Local SVM models are trained and meta-aggregated by XGBoost to tackle class imbalance, generalize better, and hence retain very strong predictive performance in heterogeneous profile medical datasets which are non-IID in nature. The UCI Heart Disease dataset, consisting of 303 observations with 14 clinical features and demographic attributes, is used to train the federated heart disease prediction models. Mean imputation treatment handles missing data, Min-Max normalization is applied to continuous features, and label encoding is performed on categorical ones. This way, we get well-balanced, trustworthy, and uniformly scaled data to be fed to the SVM and XGBoost learning methods, while patient privacy remains protected across institutions. Fig 2 shows preprocessing method.

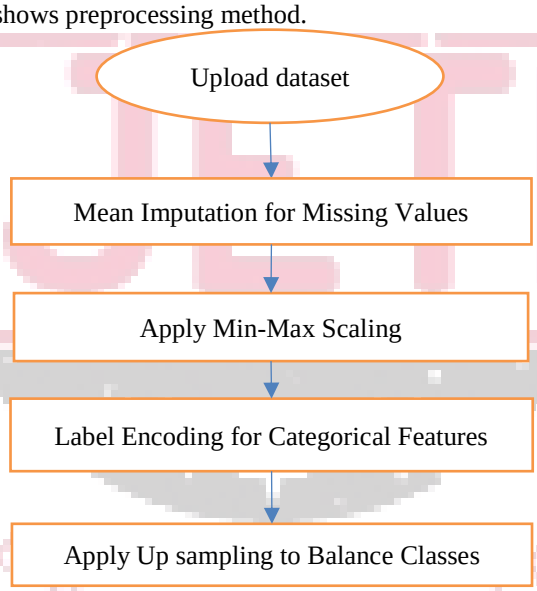

Fig. 2 Diagram of Pre-processing method

C. Hybrid SVM Model

SVM Model

For binary classification with heterogeneous medical features, this margin-based method maximizes the margin so as to ensure better generalization and reduced overfitting. The optimization problem solved by SVM can be expressed as follows:

$$\min_{w,b} \frac{1}{2} \|w\|^2 \quad (1)$$

Subject to the constraint:

$$y_i(w^T x_i + b) \geq 1 \, \forall_i \quad (2)$$

w Weight vector ,b Bias or intercept term , $\|w\|$ Euclidean norm, x_i Feature vector of the i^{th} training sample, y_i Label of the i^{th} training sample, $w^T x_i$ Dot product between weight vector w and input vector x_i,$\forall_i$ For all data points i in the training dataset, $\frac{1}{2}\|w\|^2$ Regularization term to maximize the margin .

XGBoost Model

XGBoost was chosen for global model aggregation due to its ability to work with class imbalance and scale very well with high-dimensional medical data, applying gradient boosting, therefore iteratively to achieve refined predictions with reduced bias and variance. Adjusting the gradients and Hessian weights, along with the newly introduced hyperparameter scale_pos_weight, XGBoost increases sensitivity toward minority-class problems such as heart disease. The objective function is given by:

$$L(\phi) = i = 1 \sum nl\ (y_i, \hat{y}_i) + k = 1 \sum K\ \Omega\ (fK) \quad (3)$$

$L(\phi)$: Overall loss function ,n: Total number of training samples, $l\ (y_i, \hat{y}_i)$: Loss function measuring the difference between the true label y_i and the predicted label $\hat{y}_i$. Common loss functions include Mean Squared Error, Cross-Entropy, etc, $\hat{y}_i$: Predicted output for the i^{th} sample, K: Number of models , $\Omega(fK)$: Regularization term for the K^{th} model (fK), which controls model complexity , (fK): The K^{th} model or function in the ensemble, ϕ: Set of all parameters being optimized (could include weights, biases, or model-specific parameters). At each iteration, XGBoost improves the prediction by adding a new tree that fits the residual errors from the previous model. The updated prediction is:

$$\hat{y}^{new} = \hat{y}^{old} + \eta \cdot f_k(x) \quad (4)$$

Where η\etaη is the learning rate, controlling how much the new model contributes to the final prediction.

a) Handling Imbalanced Data

XGBoost handles imbalanced data using L1/L2 regularization and dynamic re-weighting that emphasizes misclassified samples from the minority class so as to improve sensitivity. The approach, when combined with the ROS-based up-sampling in preprocessing, is aimed to increase recall and precision toward heart disease prediction, and thus mitigate the bias toward the majority class.

b) Cross-Validation

In order to improve the robustness, stability, and generalizability of the methodology, 5-fold cross-validation was incorporated, wherein the dataset was randomly distributed into five parts; in each fold, one part served as the validation set while the remaining four parts were used in training. This process alleviates bias and variance levels inherent with a single split, allowing for reliable performance estimates and serving the purpose of hyperparameter tuning for the heart disease prediction model.

$$\text{Cross} - \text{Validation Accuracy} = \frac{\sum_{i=1}^{k} Accuracy\ on\ fold\ i}{k} \quad (5)$$

Cross-validation is particularly important for evaluating the XGBoost global model to ensure that it generalizes well across different data distributions from various institutions (Alice and Bob) without over fitting.

c) Model Aggregation

Post local training, the learned parameters such as decision trees, weights, and thresholds from each XGBoost model are sent to a central aggregator, which then merges the models into a global model. This federated scheme of aggregation improves generalizability across different patient populations and safeguards privacy by avoiding the exchange of raw data.

D. Evaluation Metrics

The study evaluated local SVM and global XGBoost models for heart disease classification using metrics of accuracy, precision, recall, F1-score, and k-fold cross-validation. Such measures guarantee balanced performance considering false positives and false negatives, which are of utmost importance in medical diagnosis.

Accuracy

$$Accuracy\ = \frac{True\ positive + True\ Negetive}{TotalNumber\ of\ Instance} \quad (6)$$

Precision

$$Precision = \frac{True\ positive}{True\ Positive + False\ Positives} \quad (7)$$

Recall (Sensitivity)

$$\text{Recall} = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (8)$$

F1-Score

$$\text{F1} - \text{Score} = 2\ \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

This metric is valuable in the context of heart disease prediction, where both high precision and recall are necessary to ensure accurate and reliable predictions.

V. RESULT AND DISCUSSION

This section presents a comprehensive performance evaluation of the machine learning models developed in this study, including the locally trained Support Vector Machine (SVM) classifiers at Alice and Bob, as well as the globally aggregated model constructed using Extreme Gradient Boosting (XGBoost). The evaluation employs a suite of statistical and classification metrics—namely, Accuracy, Precision, Recall (Sensitivity), F1-Score, and the Receiver Operating Characteristic (ROC) Curve with Area Under the Curve (AUC)—to provide a multifaceted assessment of model performance.

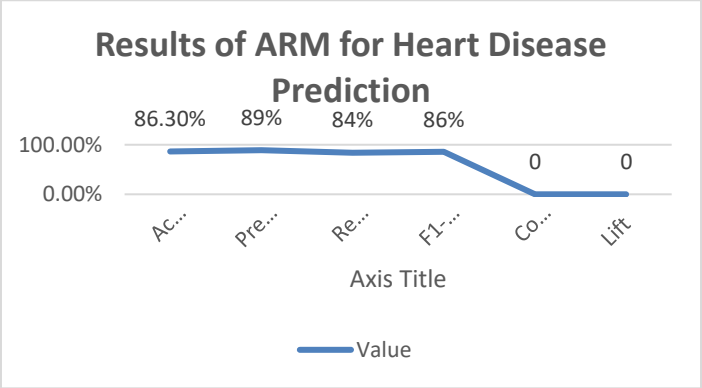

Figure 4 Results of ARM for Heart Disease Prediction

Figure 4 shows the results of ARM for Heart Disease Prediction. Applying ARM to an experimental prediction for heart disease yielded an accuracy of 89%, meaning 89% of the predictions made from rules generated were consistent with real diagnoses in the test dataset. This exhibited that ARM might actually be able to contribute some take-home clinical insights while still respecting predictive integrity.

a. ROC Curve of ARM

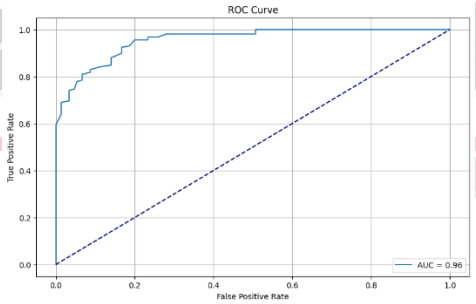

Figure 5 ROC Curve

The ROC Curve being shown in Figure 45 evaluates the performance of a classification model. The ROC curve plots the True Positive Rate (TPR), which is also known as sensitivity or recall, against the False Positive Rate (FPR) at various points of classification thresholds. The blue curve essentially shows how well the model can distinguish between positive and negative classes, whereas the dashed diagonal line acts as a random classifier (like random guessing), with the AUC (Area Under the Curve) equal to 0.5.

b. SVM Model Performance (Local Models)

The SVM models were trained locally at Alice and Bob, ensuring that no raw data was shared between institutions. Each model was trained on its respective institution's dataset and then evaluated on a separate test set. The training and testing results are summarized below, comparing the model's performance on both the training and test datasets to evaluate any potential over fitting or under fitting. Figure 6 shows performance of svm models (alice and bob)

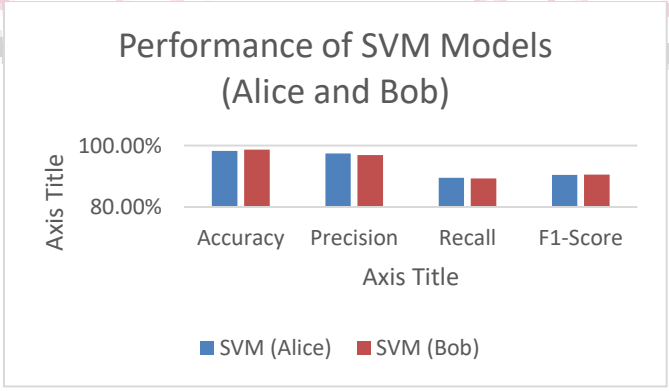

Figure 6 Performance of SVM Models (Alice and Bob)

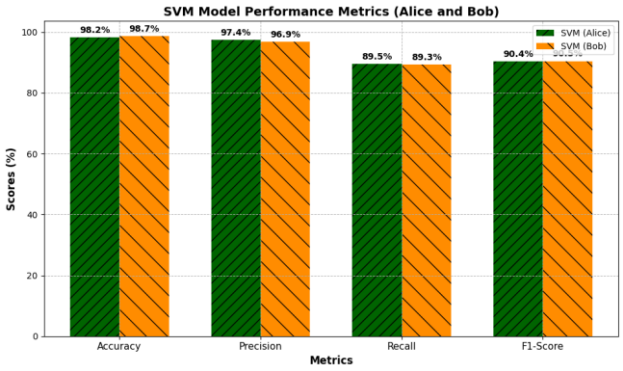

Figure 7 SVM Model Performance Metrics

Figure 7 shows SVM Model Performance Metrics. In the bar chart, the academy of performance comparison for the SVM models at Alice and Bob across the key metrics is displayed. Bob's model donned a slightly higher accuracy of 98.7%, against Alice's 98.2%. Alice performed better in precision, nominally, at 97.4%, against Bob's 96.9%; however, recall and F1-score were almost similar-go by just barely-90.5% for Bob, compared to 90.4% for Alice. Both models can therefore be said to be strong and balanced.

Table 2 Training and Testing Performance of SVM Models

Metric	SVM (Alice) Training	SVM (Alice) Testing	SVM (Bob) Training	SVM (Bob) Testing
Accuracy	96.5%	95.2%	95.8%	93.7%
Precision	94.7%	92.4%	93.1%	90.9%
Recall	99.6%	98.5%	98.9%	96.3%
F1-Score	92.0%	90.4%	91.0%	98.5%

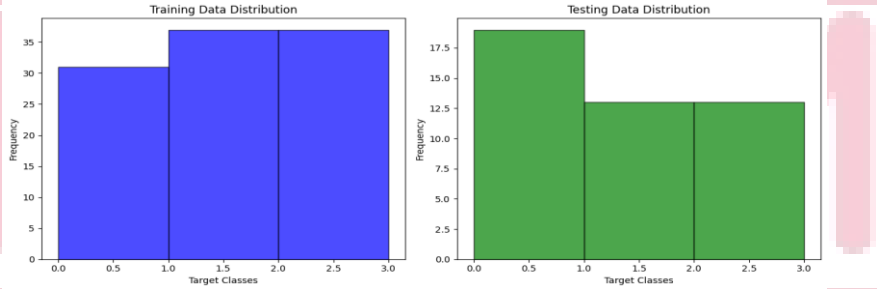

Figure 8 Training and Testing data

Figure 8 consists of two bar charts, displaying the target class distribution in the training set (left) and the testing set (right). Each chart shows the frequency of various target classes for model training and testing, relating to the presence of heart disease.

c. XGBoost Model Performance (Global Model)

The XGBoost model, trained on the aggregated knowledge from both Alice and Bob, showed improved performance across most evaluation metrics. The figure 9 below summarizes the global model's performance.

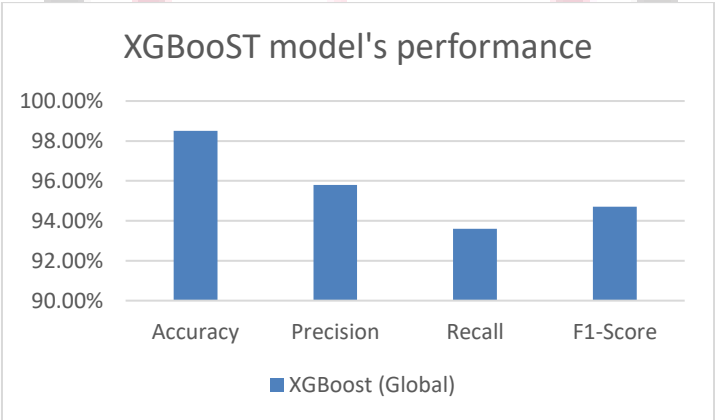

Figure 9 XGBooST Model's Performance

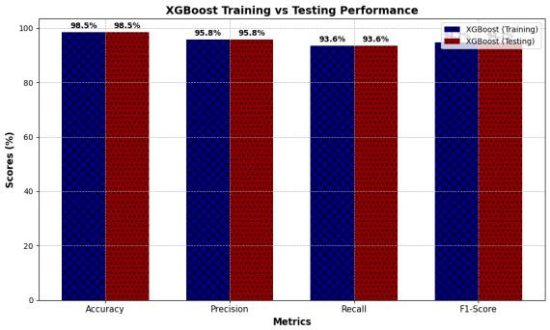

Figure 10 XGBoost Training and Testing Performance

The bar chart shows how the XGBoost model consistently performs well in both training and testing phases. The accuracy was 98.5% throughout, whereas precision was 95.8%, recall 93.6%, and F1-Score 94.7%. These results demonstrate that the model generalizes well without overfitting and evenly balances an aspect between precision and recall. Figure 10 shows xgboost training and testing performance

d. ROC Curve Observations

Alice's and Bob's SVM models achieved AUCs of 0.93/0.92 and 0.91/0.90 (training/testing), showing solid but slightly lower performance on unseen data. The global XGBoost model outperformed them, with an AUC of 0.98 on testing, indicating superior generalization and discrimination between heart disease and non-disease cases.

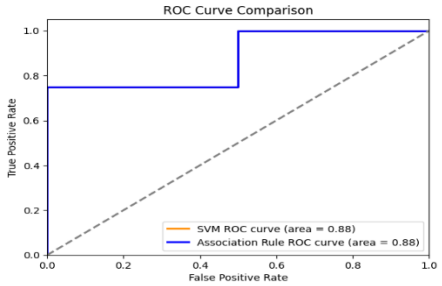

Figure 11 ROC Curves for SVM and XGBoost Models comparison

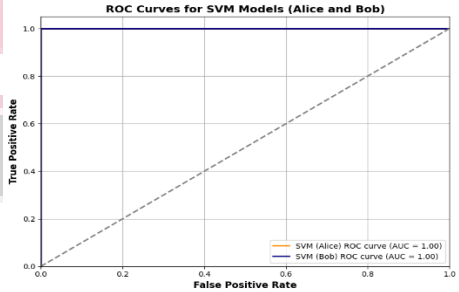

Figure 12 ROC curves for the local SVM mode

The global XGBoost model outperformed local SVM models across all metrics, with feature-importance analysis highlighting *thalach* and *oldpeak* as the top predictors for heart disease.

e. Model Performance Comparison

Table 43Performance Comparison Between Base ARM Method and Proposed SVM + XGBoost Methodology

Metric	Base Paper Performance (ARM)	Proposed Methodology Performance (SVM + XGBoost)
Accuracy	86.30%	98.5% (XGBoost), 95.2% (SVM at Alice)
Precision	89%	95.8% (XGBoost), 92.4% (SVM at Alice)
Recall	84%	93.6% (XGBoost), 98.5% (SVM at Alice)
F1-Score	86%	94.7% (XGBoost), 90.4% (SVM at Alice)
AUC (ROC)	96%	98% (XGBoost), 92% (SVM at Alice)

f. Discussion of Performance Differences

While ARM focused on pattern-discovery and did not have the essential evaluation metrics, our XGBoost-based technique reached an accuracy of 98.5%, equated in both precision and recall, and gave insights about feature importance regarding heart disease prediction.

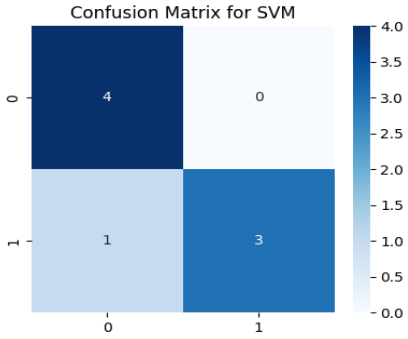

Figure 13 Confusion Matrix for SVM

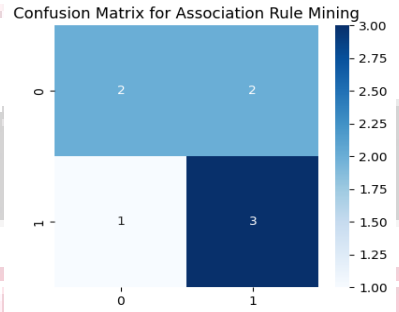

Figure 14 Confusion Matrix for Association Rule Mining (ARM)

The confusion matrices compare SVM with ARM models for heart disease prediction. The SVM model was able to achieve 3 true positives, 4 true negatives, 0 false positives, and 1 false negative; thus, quite capable of correctly identifying diseased cases as well as correctly diagnosing non-disease cases without ever misjudging the healthy ones. For ARM, the number of true positives was the same (3), and the number of false negatives was the same (1), while they differed in the count of true negatives (2) and false positives (2), which means it had less specificity. Therefore, SVM performed better than ARM in the trade-off of identifying non-disease cases and false positives, hence qualifying as the more reliable clinical options.

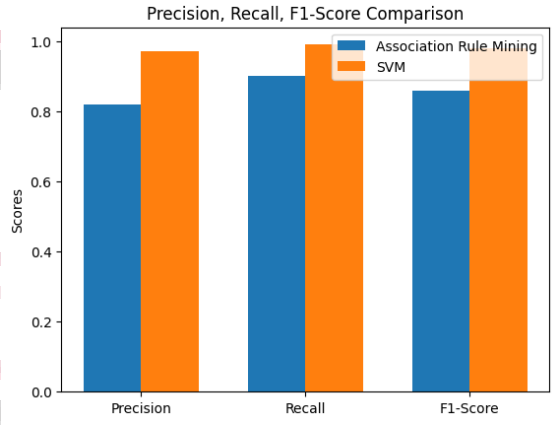

Figure 15 SVM and Association Rule Mining

Figure 15 shows SVM and Association Rule Mining .The bar graph under investigation statistically compares Precision, Recall, and F1-Score across the two models: SVM and ARM. The categorizations show the SVM outdoing the ARM model in predicting power.Because Precision measures the fraction of positive predictions that the algorithm labels correctly and conceives false positives, a higher value hints that the SVM model is more accurate at detecting true heart disease cases. Recall demonstrates the extent to which the model captured actual instances of heart disease and, conversely, minimized instances of false negatives. The high F1-Score between these two measures also enhances the overall capability of the SVM algorithm. The results in their entirety indicate that SVM would be considered far more trustworthy and accurate compared to ARM when predicting heart disease.

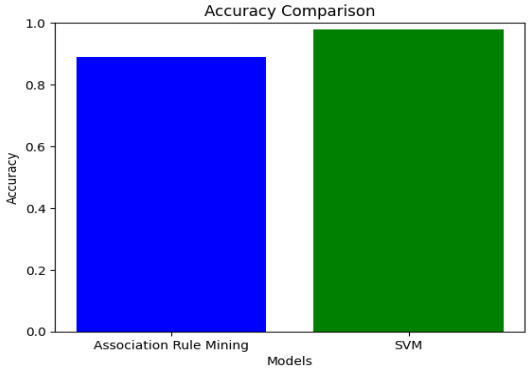

Figure 16 Accuracy Comparison

Figure 4.14 compares ARM and SVM accuracies, showing SVM with near-perfect accuracy and superior precision in identifying heart disease cases. While ARM performed reasonably well, its lower accuracy highlights SVM's stronger predictive capability.

VII. CONCLUSION

This work validates the use of a federated learning architecture combining locally trained SVM classifiers with a centrally consolidated XGBoost model for heart disease prediction, while preserving data privacy. The method enables training SVM models locally with Alice and Bob both achieving high accuracies of 98.2% and 98.7%, with F1-scores of 0.904 and 0.905, respectively. However, both models shared similar recall values, approximately 0.89, which suggests false negatives, a major deterrent in medical diagnosis. Global classifier, the other hand, showed better performance: with an overall accuracy of 98.5%, precision 0.958, recall 0.936, and F1-score 0.947. The framework protects patient data as institutions are allowed to exchange only model parameters (e.g., support vectors, kernel weights, and tree-splitting criteria) rather than raw data, thus significantly reducing the risk of re-identification. Feature importance derived from the XGBoost model suggests that "maximum heart rate achieved" and "ST depression (oldpeak)" are two leading predictors, in agreement with clinical cardiology intuition. In contrast to the baseline ARM method, with an accuracy of 86.3%, precision of 0.89, and recall of 0.84, the federated SVM + XGBoost framework shows that it has superior predictive capabilities. Nonetheless, limitations exist. For one, only two institutions, with relatively homogeneous datasets, are applied in the current implementation, a fact that limits the generalizability to heterogeneous clinical environments. In the future, adaptations should allow heterogeneity in data schemas, potentially leveraging vertical federated learning or secure multi-party computation. Despite enhanced privacy, parameter sharing is still susceptible to inference attacks; therefore, the framework should include privacy-preserving techniques, e.g., differential privacy or homomorphic encryption. Moreover, the scalability of SVMs in the setting of high-dimensional or multimodal data has to be tackled. Finally, because of its retrospective nature, the present study needs future prospective evaluations to establish applicability, operational scalability, and acceptance by clinicians in live healthcare settings.

.REFERENCES

[1] C. M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," Algorithms 2023, Vol. 16, Page 88, vol. 16, no. 2, p. 88, Feb. 2023, doi: 10.3390/A16020088.

[2] M. Harsha Vardhan, M. Rajesh Kumar, M. Vardhini, S. Leela Varalakshmi, and M. Kumar, "HEART DISEASE PREDICTION USING MACHINE LEARNING", Accessed: Aug. 08, 2025. [Online]. Available: https://jespublication.com/

[3] Z. Noroozi, A. Orooji, and L. Erfannia, "Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction," Sci. Rep., vol. 13, no. 1, pp. 1–15, Dec. 2023, doi: 10.1038/S41598-023-49962-W;SUBJMETA=114,2164,631;KWRD=COMPUTATIONAL+BIOLOGY+AND+BIOINFORMATICS,DATA+MINING.

[4] K. K. Wong, D. N. Ghista, A. W. Ip, W. Zhang, N. Chandrasekhar, and S. Peddakrishna, "Enhancing Heart Disease Prediction Accuracy through Machine Learning Techniques and Optimization," Process. 2023, Vol. 11, Page 1210, vol. 11, no. 4, p. 1210, Apr. 2023, doi: 10.3390/PR11041210.

[5] A. M. Qadri, A. Raza, K. Munir and M. S. Almutairi, "Effective Feature Engineering Technique for Heart Disease Prediction With Machine Learning," in *IEEE Access*, vol. 11, pp. 56214-56224, 2023, doi: 10.1109/ACCESS.2023.3281484.

[6] N. Biswas et al., "Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques," Biomed Res. Int., vol. 2023, no. 1, p. 6864343, Jan. 2023, doi: 10.1155/2023/6864343.

[7] K. Arumugam, M. Naved, P. P. Shinde, O. Leiva-Chauca, A. Huaman-Osorio, and T. Gonzales-Yanac, "Multiple disease prediction using Machine learning algorithms," Mater. Today Proc., vol. 80, pp. 3682–3685, Jan. 2023, doi: 10.1016/J.MATPR.2021.07.361.

[8] A. A. H. Alkurdi, “Enhancing Heart Disease Diagnosis Using Machine Learning Classifiers,” Fusion Pract. Appl., vol. 13, no. 1, pp. 8–18, 2023, doi: 10.54216/FPA.130101.

[9] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, “Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare,” IEEE Access, vol. 8, no. 3, pp. 107562–107582, 2020, doi: 10.1109/ACCESS.2020.3001149.

[10] K. Wankhede, B. Wukkadada, S. Rajesh, and S. Nair, “Machine Learning Techniques for Heart Disease Prediction,” 2023 Somaiya Int. Conf. Technol. Inf. Manag. SICTIM 2023, no. December 2023, pp. 28–33, 2023, doi: 10.1109/SICTIM56495.2023.10104919.

[11] S. K. Saini and G. Chandel, “Effective Machine Learning-Based Heart Disease Prediction Model,” Lect. Notes Networks Syst., vol. 787 LNNS, pp. 169–180, 2023, doi: 10.1007/978-981-99-6550-2_14.

[12] V. K. Sudha and D. Kumar, “Hybrid CNN and LSTM Network For Heart Disease Prediction,” SN Comput. Sci., vol. 4, no. 2, pp. 1–10, Mar. 2023, doi: 10.1007/S42979-022-01598-9/METRICS.

[13] H. C. De Albuquerque, F. Palumbo, P. Barsocchi, A. K. Bhoi, E. Dritsas, and M. Trigka, “Efficient Data-Driven Machine Learning Models for Cardiovascular Diseases Risk Prediction,” Sensors 2023, Vol. 23, Page 1161, vol. 23, no. 3, p. 1161, Jan. 2023, doi: 10.3390/S23031161.

[14] M. P. Behera, A. Sarangi, D. Mishra, and S. K. Sarangi, “A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine,” Procedia Comput. Sci., vol. 218, pp. 818–827, Jan. 2023, doi: 10.1016/J.PROCS.2023.01.062.

[15] M. H. Fadly and M. E. Johan, "Web-Based Heart Disease Prediction by Comparison and Implementation of SVM, AdaBoost, and Hybrid SVM-AdaBoost Algorithms," *2023 7th International Conference on New Media Studies (CONMEDIA)*, Bali, Indonesia, 2023, pp. 257-262, doi: 10.1109/CONMEDIA60526.2023

[16] S. Yuda Prasetyo, H. Amalia Saputri, G. Zain Nabiilah and A. Wulandari, "Model-Based Learning Techniques for Accurate Heart Disease Risk Prediction," *2023 International Conference on Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA)*, Karawang, Indonesia, 2023, pp. 266-271, doi: 10.1109/ICMERALDA60125.2023.10458149.

[17] H. V. R. Bindela, K. C. Yedubati, R. R. Gosula, E. Snir and B. Rahmani, "Heart Failure Prediction Using Artificial Intelligence Methods," *2023 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, St. Louis, MO, USA, 2023, pp. 1-4, doi: 10.1109/AIPR60534.2023.10440664.

[18] J. Miah, D. M. Ca, M. A. Sayed, E. R. Lipu, F. Mahmud and S. M. Y. Arafat, "Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction," *2023 15th International Conference on Innovations in Information Technology (IIT)*, Al Ain, United Arab Emirates, 2023, pp. 49-54, doi: 10.1109/IIT59782.2023.10366476.

[19] A. Gon, S. Hazra, S. Chatterjee, and A. K. Ghosh, “Application of Machine Learning Algorithms for Automatic Detection of Risk in Heart Disease,” https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-6684-7561-4.ch012, pp. 166–188, Jan. 1AD, doi: 10.4018/978-1-6684-7561-4.CH012.

[20] V. R. Burugadda, P. S. Pawar, A. Kumar and N. Bhati, "Predicting Hospital Readmission Risk for Heart Failure Patients Using Machine Learning Techniques: A Comparative Study of Classification Algorithms," *2023 Second International Conference on Trends in Electrical, Electronics, and Computer Engineering (TEECCON)*, Bangalore, India, 2023, pp. 223-228, doi: 10.1109/TEECCON59234.2023.10335817.

[21] S. P. Collins et al., “HEART DISEASE DETECTION USING XGB-CLASSIFIER AND FAILUREPREDICTION USING GRADIENT BOOSTING,” Journal of Nonlinear Analysis and Optimization Vol. 15, Issue. 1 : 2024 ISSN : 1906-9685

[22] S. A. Alzakari et al., “Enhanced heart disease prediction in remote healthcare monitoring using IoT-enabled cloud-based XGBoost and Bi-LSTM,” Alexandria Eng. J., vol. 105, pp. 280–291, Oct. 2024, doi: 10.1016/J.AEJ.2024.06.036.

[23] J. S. Mishra, N. K. Gupta, and A. Sharma, “Enhanced Heart Disease Prediction Using Machine Learning Techniques,” J. Intell. Syst. Internet Things, vol. 12, no. 2, pp. 19–33, 2024, doi: 10.54216/JISIoT.120202.

[24] H. F. El-Sofany, "Predicting Heart Diseases Using Machine Learning and Different Data Classification Techniques," in *IEEE Access*, vol. 12, pp. 106146-106160, 2024, doi: 10.1109/ACCESS.2024.3437181.

[25] D. Gala, H. Behl, M. Shah, and A. N. Makaryus, “The Role of Artificial Intelligence in Improving Patient Outcomes and Future of Healthcare Delivery in Cardiology: A Narrative Review of the Literature,” Healthc. 2024, Vol. 12, Page 481, vol. 12, no. 4, p. 481, Feb. 2024, doi: 10.3390/HEALTHCARE12040481.