AI-based federated learning for heart disease prediction: a collaborative and privacy-preserving approach

Stuti Bhatt[1], Surender Reddy Salkuti[2], Seong-Cheol Kim[2]
[1]Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun, India
[2]Department of Railroad and Electrical Engineering, Woosong University, Daejeon, Republic of Korea

Article Info

Article history:

Received Oct 22, 2024
Revised Jan 26, 2025
Accepted Jun 9, 2025

Keywords:

Disease prediction
Distributed training
Federated learning
Heart disease
Machine learning

ABSTRACT

People with symptoms like diabetes, high BP, and high cholesterol are at an increased risk for heart disease and stroke as they get older. To mitigate this threat, predictive fashions leveraging machine learning (ML) and artificial intelligence (AI) have emerged as a precious gear; however, heart disease prediction is a complicated task, and diagnosis outcomes are hardly ever accurate. Currently, the existing ML tech says it is necessary to have data in certain centralized locations to detect heart disease, as data can be found centrally and is easily accessible. This review introduces federated learning (FL) to answer data privacy challenges in heart disease prediction. FL, a collaborative technique pioneered by Google, trains algorithms across independent sessions using local datasets. This paper investigates recent ML methods and databases for predicting cardiovascular disease (heart attack). Previous research explores algorithms like region-based convolutional neural network (RCNN), convolutional neural network (CNN), and federated logistic regressions (FLRs) for heart and other disease prediction. FL allows the training of a collaborative model while keeping patient info spread out among various sites, ensuring privacy and security. This paper explores the efficacy of FL, a collaborative technique, in enhancing the accuracy of cardiovascular disease (CVD) prediction models while preserving data privacy across distributed datasets.

Corresponding Author:

Surender Reddy Salkuti
Department of Railroad and Electrical Engineering, Woosong University
Jayang-Dong, Dong-Gu, Daejeon - 34606, Republic of Korea
Email: surender@wsu.ac.kr

1. INTRODUCTION

A heart attack is a scientific urgency wherein the patient's coronary heart muscle starts to die as the blood flow to a section of the heart is obstructed, it results in damage or death of the heart muscle tissue. A blockage inside the arteries that supply blood to your heart commonly causes this. If a medical practitioner is not able to repair blood flow as soon as possible, a coronary heart attack can cause permanent heart damage which results in the death of the patient. A myocardial infarction is a hazardous circumstance that takes place due to the fact you don't have sufficient blood flow to a number of your heart muscles. It's expected around 2 hundred million humans are living with coronary heart disease. Globally around a hundred and ten million guys and eighty million ladies have coronary heart disease. The present-day occurrence of diabetes mellitus is 463 million, equal to 9.3% of the world population. The international pandemic of diabetes is predicted to elevate this figure to 578 million (10.2%) through the year 2030 and seven hundred million (10.9%) through 2045.

Federated learning (FL) can play a significant role in addressing challenges associated with coronary heart attacks (myocardial infarction) and diabetes mellitus (DM) through leveraging collaborative records sharing at the same time as preserving privacy and security. Here's how FL may be useful in those contexts:

- Privacy-preserving model training: FL permits the training of the machine learning (ML) models throughout decentralized records sources (together with medical statistics from different healthcare providers) without sharing raw records. This is important for preserving patient privacy, as sensitive health information like cardiac health records and diabetes status can remain localized [1].
- Improved model generalization: by leveraging numerous records from various sources (reflecting special demographics, genetic backgrounds, and medical histories), FL can assist develop extra sturdy and generalizable models for predicting and preventing coronary heart attacks and diabetes complications.
- Personalized medicine: FL permits the improvement of personalized predictive models for assessing individual dangers of coronary heart attacks or diabetes-associated complications primarily based totally on diverse patient records. This can result in extra-centered interventions and remedy strategies [2].
- Real-time monitoring and alerts: FL models can be constantly updated and refined through the usage of real-time records from a couple of assets, permitting faster detection of early warning signs related to coronary heart attacks or diabetes exacerbations. This can facilitate timely interventions and preventive measures.
- Scaling and collaboration: given the worldwide incidence of those conditions, FL provides a scalable method for collaboration throughout institutions, regions, and countries. This collective attempt can result in a higher understanding of sickness patterns, risk factors, and powerful control strategies.
- Integration with wearable devices and internet of things (IoT): FL can combine records from wearable devices and IoT devices used for non-stop fitness monitoring. This complete records aggregation can enhance predictive models for figuring out pre-coronary heart attack signs and symptoms or diabetic complications.

A diabetic person has a higher risk of having a coronary heart attack because of attributes like insulin resistance and chronic excessive blood sugar levels this is why it's far essential that ailments like diabetes, blood pressure, and cholesterol degree are defined before and given more attention as they have got better probabilities of hazard. Sharma and Sharma [1] highlights the use of FL to deal with data privacy in heart disease prediction, accomplishing comparable accuracy to centralized models, and indicates future exploration of transfer learning for stronger predictive performance. Bharathi *et al.* [2] examines leveraging FL for coronary artery disease prediction using federated logistic regression (FLR) and federated support vector machine (SVM) and also states that it ensures data privacy, reaching as much as 95.8% accuracy, for FLR centralized models, through local statistics aggregation. Hayyolalam *et al.* [3] discusses the use of edge computing and a hybrid ML technique with black widow optimization for heart disease prediction, attaining 90.11% accuracy, and suggests future integration with FL. Sharma and Sharma [4] examines FL for heart disease detection, reaching 94.99% accuracy with a convolutional neural network (CNN) model, emphasizing privacy protection, and suggesting future exploration of different ML and deep learning (DL) techniques. Yuan *et al.* [5] explores a virtual twin-assisted FL protocol for disease prediction, which will improve performance as well as accuracy. Future studies aim to optimize transfer learning for improved medical image recognition. Dolo *et al.* [6] examines the A model for predicting diabetes using differentially private stochastic gradient descent federated averaging (DPSGDFedAvg), reaching 60-70% accuracy at the same time as making sure statistics is private. Future research objectives are to enhance model accuracy and privacy. Khan *et al.* [7] uses FL for disease prediction in unprecedented areas using chest X-rays, achieving a 2% efficiency improvement and an area under the curve (AUC) of 77.91, while preserving patient privacy. Future research will broaden scan types and enhance dataset diversity. Nandhini *et al.* [8] investigates FL for chronic kidney disease prediction, highlighting improved accuracy and efficiency through image processing and decentralized data training. Future research aims to enhance prediction by refining algorithm performance and addressing age-related disease progression.

This paper presents a comprehensive and critical overview of the current state of research in this area and also categorizes research based on different approaches, such as types of FL algorithms used, privacy-preserving techniques, and specific applications in heart disease prediction. The paper also identifies the research gaps and limitations. The overview assesses the impact of federated learning on coronary heart disease prediction and overall healthcare practices, highlighting case research and practical implementations.

2. WORKING

ML approaches commonly require centralizing the training data into a common store and it is depicted in Figure 1. The limitation of this centralized data is the communication exchange between the server and the client as it could be time-consuming and hurt user experience due to network latency, battery life, and connectivity. FL (collaborative learning) is a sub-field of ML and is a decentralized approach to training the ML models where user data is never transmitted to a central server.

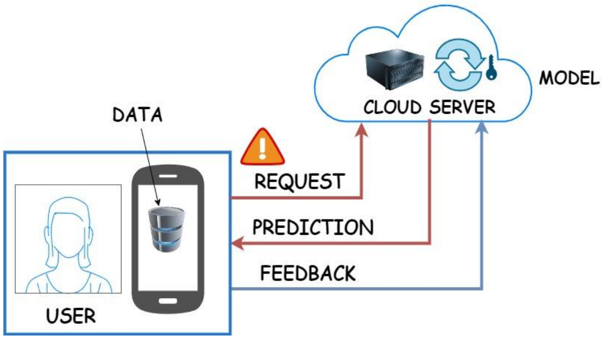

Figure 1. Centralized system

Start with distributing the model from the server to the client but deployment to every client must be strategic to avoid interruptions in user experience. Hence our primary task is to identify the clients that are available and plugged in, and not in active use. Additionally, discover which ones are suitable and possess sufficient data, as not all clients may meet this criterion. Once suitable devices (user1, user2, and user3) are identified model can be deployed. Each client then independently trains its model locally using their own data at their end (locally) generating a new model locally, which is further sent to the server. The point here is the data used to train the model remains on respective devices and is not transmitted. Only the model parameters, such as weights (w1, w2, and w3), and biases, are sent to the server. The server processes all the locally trained models and then performs aggregation, effectively creating a new model. To know if this process is making a good meaningful model by doing the process repeatedly and with every round, the combined model gets a little bit better with the help of data gained from all the clients. For additional security in FL, we can use the concept of secure aggregation, where the server pairs up devices with each other in a buddy system. The federated system is depicted in Figure 2.

Organizations like hospitals can also be regarded as remote or local devices that include a large number of patient data for predictive healthcare [9], [10]. However, hospitals perform under strict privacy practices and might face legal, administrative, or moral constraints that require data to remain local. FL is a promising answer for these applications, as it may reduce pressure on the network and allow private learning among numerous corporations. We found that compared to other methodologies used for disease prediction federated learning gives better result whether it is about privacy, security management or better result and performance, FL techniques ace them all.

As FL simplifies everything it also has certain limitations, one of them is the problem formulation. The typical FL problem involves studying as well as developing a single, global statistical model using data stored on tens to probably hundreds of thousands of remote devices. In particular, the aim is usually to minimize the subsequent objective function:

$$\frac{min}{\omega}\ F(\omega),\ \ where\ F(\omega) = \sum_{k=1}^{m} P_k F_k(\omega) \tag{1}$$

where m represents the total number of devices, F_k denotes local objective function for the kth device, and P_k indicates the relative impact of each device with $P_k \geq 0$.

$$\sum_{k=1}^{m} P_k = 1 \tag{2}$$

The local objective function F_k is typically defined as the empirical risk calculated from the local data. The user determines the relative impact of each device P_k, commonly set to P_k= 1/m or P_k= nk/n, in this context n represents the total number of samples across all devices [11], [12]. While this is a common objective in FL, alternative approaches exist, such as concurrently learning related local models through multi-task learning where each and every device corresponds to a distinct task. Both the multi-task and meta-learning perspectives offer avenues for personalized or device-specific modeling, effectively addressing the statistical heterogeneity of the data.

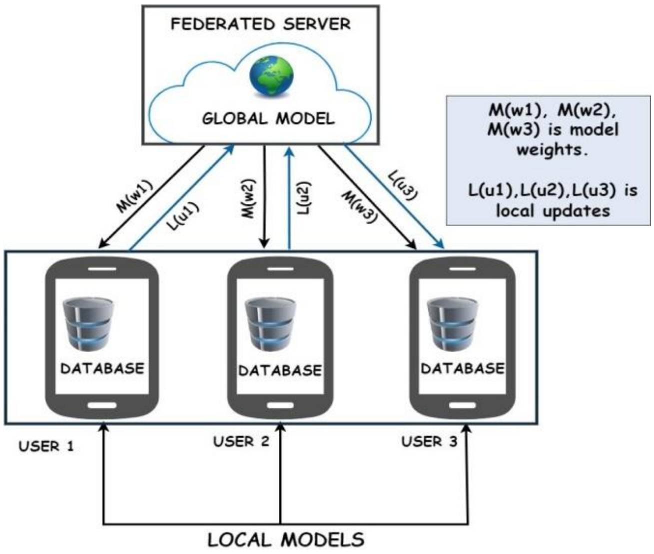

Figure 2. Federated systems

3. FEDERATED LEARNING IN HEALTHCARE

Organizations such as hospitals can be viewed as remote devices that house extensive patient records for predictive healthcare purposes [13]. However, hospitals function under strict privacy protocols and may encounter legal, administrative, or moral/ethical constraints that require data locality. FL emerges as a promising solution for such scenarios, as it may alleviate stress on the network enabling confidential learning among multiple devices or organizations. Figure 3 illustrates an application scenario wherein a model is trained from distributed digital health records. The local data consists of the patient's history or medical records which remains confidential [14]. Privacy and security are a major concern of any organization. Unlike conventional centralized system learning, wherein data is collected and processed in a central server, FL allows models to be trained throughout more than one decentralized server while preserving the statistics localized. Minimizing the transfer of statistics drastically lowers the probability of data interception or leakage [15]. FL permits the aggregation of knowledge from numerous datasets throughout special institutions, leading to extra strong and generalized models.

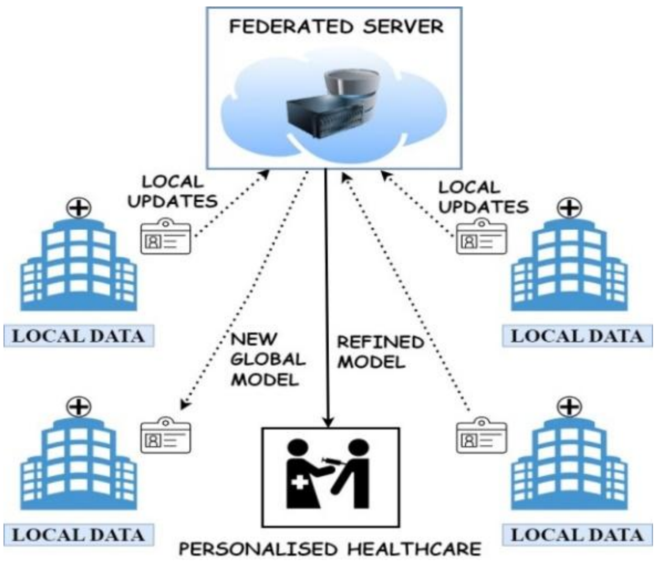

Figure 3. FL in healthcare

4. ALGORITHMS

ML and information mining-based techniques for the prediction and detection of coronary heart disorder could be of superb scientific utility however are relatively challenging to develop. In most countries, there's a lack of cardiovascular knowledge and an enormous price of incorrectly identified instances which can be addressed by developing correct and efficient early-stage coronary heart disorder prediction by analytical guide of clinical decision-making with virtual patient records [16], [17]. This observation aimed to discover ML classifiers with the highest accuracy for such diagnostic purposes. Several supervised ML algorithms were carried out and compared for overall performance and accuracy in coronary heart disease prediction. Some major ML algorithms are discussed in the following section:

– Support vector machine (SVM): a type of supervised learning used for classification and regression [18], [19]. The main idea behind developing is to find the hyperplane in a high-dimensional space that maximally separates the different classes which results in précised classification. For linear SVM classifier:

$$y = \begin{cases} 1: \omega^T x + b \geq 0 \\ 0: \omega^T x + b < 0 \end{cases} \tag{3}$$

– Logistic regression (LR): used to predict bi-classification problems, such as email spam or not, yes or no, 0 or 1, true or false. LR uses a sigmoid function, i.e., a logistic function [14].

$$f(x) = \frac{1}{1+e^{-x}} \tag{4}$$

Sigmoid function is simply trying to convert the independent variable (x) into an expression of probability that ranges between 0 and 1 with respect to the dependent variable. LR with multiple independent variables:

$$y = f(\beta 0 + \beta 1 \times 1 + \beta 2 \times 2 + \ \dots\dots\dots \beta n \times n) \tag{5}$$

where β is the regression coefficient.

– Linear regression: modelling relationships are linear in nature. It is a facts evaluation approach that predicts the value of unknown facts by using another associated and known data value. Like if the value of x in increasing value of y will also increase [20]. Simple linear regression:

$$y = \alpha 0 + \alpha 1 + x1 \tag{6}$$

Where $x1$ is the independent variable, y is the dependent variable, $\alpha 0$, $\alpha 1$ are coefficients of regression. The Multiple Linear regressions are represented by,

$$y = \alpha 0 \times x0 + \alpha 1 \times x1 \ + \ \alpha 2 \times x2 + \ \dots\dots .\alpha n \times xn + \epsilon \tag{7}$$

– Random forest: type of ensemble ML model, where multiple models are working together to make a prediction. In the case of random forest, the smaller models are decision trees [21], [22]. Each individual decision tree will make a prediction and then the final prediction is made after aggregation of all the previous outputs/ predictions [18], [19].

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(fi - yi)^2 \tag{8}$$

Where MSE is a mean square error, for n number of data points, fi is the value returned by each decision tree model ensemble, i is a data point and yi is the actual value for i.

– Naïve Bayes (NB) classifier: supervised learning based on probabilistic logic uses Bayes' theorem. Here naïve part came from an assumption stating features in data are independent, (not in the case of real-world data) [23]. The probability (P) is represented by,

$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)} \tag{9}$$

5. COMPARATIVE ANALYSIS

Table 1 includes a comparative analysis between different works for the purpose of detection of cardiovascular disease (CVD). This analysis includes a summary of the methodology applied by different authors and the corresponding results obtained after the application of the proposed methodologies [24], [25]. The key findings are also separately mentioned which suggests the effectiveness of the proposed methodology. Privacy and security feature is also mentioned in the above works [26]. The comparative evaluation of artificial intelligence (AI)-primarily based FL for CVD prediction suggests that FL can achieve accuracy near a centralized model while maintaining information privacy. For example, federated LR and federated SVM accomplished as much as 95.8 accuracy and federated CNN model reached 94.99 accuracy, nearing the centralized fashion of 97% [27]. Overall, FL affords a strong, privacy-maintaining opportunity compared to centralized approaches, with future studies having to optimize its effectiveness in healthcare.

Table 1. Comparative analysis between different works for the purpose of detection of CVD

Methodology	Key findings	Algorithms	Data source	P&S
FL using linear regression and SVM [2]	Predicting heart disease accuracy: FL= 89%, centralized = 95.8; FSVM performed well compared to FLR.	FLR, federated support vector machine (FSVM)	UCI Machine learning repository.	✓
Random oversampling on the dataset including samples of about 1 lakh. FL with CNN on UCI Cleveland dataset [4]	For heart disease prediction accuracy: federated CNN = 94.99%, closed to centralized CNN = 97%	FL-CNN	UCI Cleveland dataset	✓
Classification techniques (decision tree (DT), linear regression, polynomial regression, NB, SVM, CNN) are compared, and missing values are predicted using single and multi-value imputations [8]	For chronic kidney disease. Accuracy: LR= 98.4 Polynomial R.: 95.3 NB= 91.3 DT= 93.6 FL= 98.7	DT, linear regression, Polynomial regression, NB and proposed FL	Kaggle dataset	✓
mRMR feature selection with LR and SVM [9]	Predicting Heart disease. Accuracy: FL= 85% Centralized=36% and similar F1-scores, FL=91%, centralized= 92%	LR SVM	PTB diagnostic ECG database	✓
Correlation-based feature selection and oversampling, SMOTE with a variety of ML algorithms on a dataset with 14 features and 1000 records [10]	Prediction of heart disease Accuracy% after sampling (oversampling, smote): LR = 96.1%, 100% NB = 94.7%,100% KNN = 97%,98.36% Random forest = 99.20%,100%, SVM = 97.60%,100%, Tree = 98.75%,100%	Random oversampling, K-nearest neighbor (KNN), and SMOTE technique	Public dataset	✓
XGBoost, AdaBoost, random forest, DT, NB, and LR. Kaggle dataset [28]	Prediction of heart disease LR model accuracy = 91.57%	XGBoost, AdaBoost, random forest, DT, LR, NB	Kaggle dataset (319,795)	✓
To handle classes that are imbalanced researchers use SMOTE, and MIN-MAX normalization is applied afterward. At the end 3 ML classifiers are used i.e. Gaussian NB, KNN for K = 5, SVM (‘rbf’) [29]	Prediction of heart disease Precision= 92.50% Recall, 92.22% and F1-score =92.36%	Gaussian, NB, SVM, KNN, soft voting	Cleveland heart disease dataset taken from UCI machine learning repository	✓
Proposed hybrid random forest with a linear model (HRFLM) [11]	Accuracy = 88.4, precision=90.1, F-measure=90, sensitivity=92.8, specificity= 82.6	ML techniques, HRFLM.	Cleveland dataset collected from a UCI ML repository.	X
Designed ACVD-HBOMDL technique with feature selection and hyperparameter tuning [12]	ACVD-HBOMDL achieved 99.39% accuracy in CVD diagnosis.	Min-max scaler for data preprocessing Honey badger optimization (HBO) algorithm for feature selection, deep learning modified neural network classifier, Bayesian optimization for hyperparameter tuning	Kaggle repository: aggregated from various regions and Statlog datasets.	✓
CV with K-fold, Data split into training, cross-validation, and testing; evaluated with K-fold CV [13]	Accuracy = 97.32%, recall= 97.58%, Precision=97.16%, F1-measure=97.37%, specificity =96.87% G-mean = 97.22%	NB, K-NN, SVM, random forest, artificial neural network (ANN)	Specific medical dataset (repository not mentioned)	✓

6. CONCLUSION

This paper explores the utility of AI-based FL for coronary heart disease prediction, emphasizing its importance in maintaining an affected person's privacy even as leveraging distributed records from hospitals and other organizations. FL emerges as a powerful solution, addressing the restrictions of traditional centralized ML models, in particular within the context of data which is very sensitive for the patient. The comparative evaluation among centralized and FL strategies exhibits that FL can gather competitive accuracy rates. For instance, the FL-CNN model validated a validation accuracy of 94.99% on the UCI Cleveland dataset, which is remarkably close to the 97% accuracy performed by centralized CNN models. This highlights FL`s capability to offer effective predictive overall performance even as ensuring that affected person information stays stable and decentralized.

ACKNOWLEDGEMENT

This research work was supported by "Woosong University's Academic Research Funding - 2025".

FUNDING INFORMATION

This research work was supported by "Woosong University's Academic Research Funding - 2025".

AUTHOR CONTRIBUTIONS STATEMENT

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Stuti Bhatt	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	
Surender Reddy Salkuti	✓	✓		✓	✓	✓	✓		✓	✓	✓	✓		✓
Seong-Cheol Kim		✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

INFORMED CONSENT

Not applicable — this study did not involve human participants requiring informed consent.

ETHICAL APPROVAL

Not applicable — this study did not involve human participants or animals.

DATA AVAILABILITY

Data availability does not apply to this article as no new data were created or analyzed in this study.

REFERENCES

[1] P. Sharma and S. Sharma, "A comprehensive study of the machine learning with federated learning approach for predicting heart disease," in *Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2023*, Sep. 2023, pp. 1867–1873, doi: 10.1109/IC3I59117.2023.10397716.

[2] S. K. Bharathi, M. Dhavamani, and K. Niranjan, "A federated learning based approach for heart disease prediction," in *Proceedings - 6th International Conference on Computing Methodologies and Communication, ICCMC 2022*, Mar. 2022, pp. 1117–1121, doi: 10.1109/ICCMC53470.2022.9754119.

[3] V. Hayyolalam, S. Otoum, and O. Ozkasap, "A hybrid edge-assisted machine learning approach for detecting heart disease," in *IEEE International Conference on Communications*, May 2022, vol. 2022-May, pp. 2966–2971, doi: 10.1109/ICC45855.2022.9838933.

[4] P. Sharma and S. Sharma, "An effective FL-CNN based data securing model for heart disease prediction," in *Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2023*, Sep. 2023, pp. 1862–1866, doi: 10.1109/IC3I59117.2023.10397649.

[5] X. Yuan, J. Zhang, J. Luo, J. Chen, Z. Shi, and M. Qin, "An efficient digital twin assisted clustered federated learning algorithm for disease prediction," in *IEEE Vehicular Technology Conference*, Jun. 2022, vol. 2022-June, pp. 1–6, doi: 10.1109/VTC2022-Spring54318.2022.9860704.

[6] B. Dolo, F. Loukil, and K. Boukadi, "Early detection of diabetes mellitus using differentially private SGD in federated learning," in *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, Dec. 2022, vol. 2022-Decem, pp. 1–8, doi: 10.1109/AICCSA56895.2022.10017908.

[7] S. Khan, L. S. Palanisamy, and M. Raghuraman, "Federated learning-a novel approach for predicting diseases in unprecented areas," in *6th International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2024*, Feb. 2024, pp. 58–63, doi: 10.1109/ICAIIC60209.2024.10463360.

[8] J. M. Nandhini, S. Joshi, and K. Anuratha, "Federated learning based prediction of chronic kidney diseases," in *2022 1st International Conference on Computational Science and Technology, ICCST 2022 - Proceedings*, Nov. 2022, pp. 851–856, doi: 10.1109/ICCST55948.2022.10040317.

[9] B. Ülver, R. A. Yurtoğlu, H. Dervişoğlu, R. Halepmollasi, and M. Haklidir, "Federated learning in predicting heart disease," in *31st IEEE Conference on Signal Processing and Communications Applications, SIU 2023*, Jul. 2023, pp. 1–4, doi: 10.1109/SIU59756.2023.10223935.

[10] F. Sabah, Y. Chen, Z. Yang, A. Raheem, M. Azam, and R. Sarwar, "Heart disease prediction with 100% accuracy, using machine learning: performance improvement with features selection and sampling," in *Proceedings of 2023 8th IEEE International Conference on Network Intelligence and Digital Content, IC-NIDC 2023*, Nov. 2023, pp. 41–45, doi: 10.1109/IC-NIDC59918.2023.10390693.

[11] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[12] M. Obayya, J. M. Alsamri, M. A. Al-Hagery, A. Mohammed, and M. A. Hamza, "Automated cardiovascular disease diagnosis using honey badger optimization with modified deep learning model," *IEEE Access*, vol. 11, pp. 64272–64281, 2023, doi: 10.1109/ACCESS.2023.3286661.

[13] C. Chakraborty and A. Kishor, "Real-time cloud-based patient-centric monitoring using computational health systems," *IEEE Transactions on Computational Social Systems*, vol. 9, no. 6, pp. 1613–1623, Dec. 2022, doi: 10.1109/TCSS.2022.3170375.

[14] A. Yashudas, D. Gupta, G. C. Prashant, A. Dua, D. Alqahtani, and A. S. K. Reddy, "DEEP-CARDIO: recommendation system for cardiovascular disease prediction using IoT network," *IEEE Sensors Journal*, vol. 24, no. 9, pp. 14539–14547, May 2024, doi: 10.1109/JSEN.2024.3373429.

[15] G. Koutitas, K. Nolen, S. Attal, A. Ventouris, Y. Dolev, and H. T. Van Den Broek, "Technical feasibility of implementing and commercializing a machine learning model for rare disease prediction," *IEEE Access*, vol. 11, pp. 84430–84439, 2023, doi: 10.1109/ACCESS.2023.3299866.

[16] B. T. H. Dang, P. H. Luan, V. D. T. Ngan, N. T. Trong, P. T. Duy, and V. H. Pham, "TrustFedHealth: federated learning with homomorphic encryption and blockchain for heart disease prediction in the smart healthcare," in *International Conference on Advanced Technologies for Communications*, Oct. 2023, pp. 178–183, doi: 10.1109/ATC58710.2023.10318944.

[17] R. Kapila, T. Ragunathan, S. Saleti, T. J. Lakshmi, and M. W. Ahmad, "Heart disease prediction using novel quine McCluskey binary classifier (QMBC)," *IEEE Access*, vol. 11, pp. 64324–64347, 2023, doi: 10.1109/ACCESS.2023.3289584.

[18] N. Mohan, V. Jain, and G. Agrawal, "Heart disease prediction using supervised machine learning algorithms," in *2021 5th International Conference on Information Systems and Computer Networks, ISCON 2021*, Oct. 2021, pp. 1–3, doi: 10.1109/ISCON52037.2021.9702314.

[19] S. J. Pasha and E. S. Mohamed, "Novel feature reduction (NFR) model with machine learning and data mining algorithms for effective disease risk prediction," *IEEE Access*, vol. 8, pp. 184087–184108, 2020, doi: 10.1109/ACCESS.2020.3028714.

[20] S. Bhatt, S. Sharma, and S. Bhadula, "Advanced blockchain applications in cryptocurrency and various sectors," in *2023 1st International Conference on Circuits, Power, and Intelligent Systems, CCPIS 2023*, Sep. 2023, pp. 1–6, doi: 10.1109/CCPIS59145.2023.10291465.

[21] G. Muhammad *et al.*, "Enhancing prognosis accuracy for ischemic cardiovascular disease using K nearest neighbor algorithm: a robust approach," *IEEE Access*, vol. 11, pp. 97879–97895, 2023, doi: 10.1109/ACCESS.2023.3312046.

[22] G. Joo, Y. Song, H. Im, and J. Park, "Clinical implication of machine learning in predicting the occurrence of cardiovascular disease using big data (Nationwide Cohort Data in Korea)," *IEEE Access*, vol. 8, pp. 157643–157653, 2020, doi: 10.1109/ACCESS.2020.3015757.

[23] D. Y. Omkari and K. Shaik, "An integrated two-layered voting (TLV) framework for coronary artery disease prediction using machine learning classifiers," *IEEE Access*, vol. 12, pp. 56275–56290, 2024, doi: 10.1109/ACCESS.2024.3389707.

[24] V. Veeramsetty, D. Rakesh Chandra, and S. R. Salkuti, "Short term active power load forecasting using machine learning with feature selection," in *Lecture Notes in Electrical Engineering*, vol. 824, Springer Nature Singapore, 2022, pp. 103–124.

[25] S. R. Salkuti, "A survey of big data and machine learning," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 1, pp. 575–580, Feb. 2020, doi: 10.11591/ijece.v10i1.pp575-580.

[26] K. Badapanda, D. P. Mishra, and S. R. Salkuti, "Agriculture data visualization and analysis using data mining techniques: application of unsupervised machine learning," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 20, no. 1, pp. 98–108, Feb. 2022, doi: 10.12928/TELKOMNIKA.v20i1.18938.

[27] D. P. Mishra, S. Mishra, S. Jena, and S. R. Salkuti, "Image classification using machine learning," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 3, pp. 1551–1558, 2023, doi: 10.11591/ijeecs.v31.i3.pp1551-1558.

[28] M. Mamun, M. M. Uddin, V. Kumar Tiwari, A. M. Islam, and A. U. Ferdous, "MLHeartDis:Can Machine Learning Techniques Enable to Predict Heart Diseases?," in *2022 IEEE 13th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2022*, Oct. 2022, pp. 561–565, doi: 10.1109/UEMCON54665.2022.9965714.

[29] R. A. Nazri, S. Das, and R. T. Haque Promi, "Heart disease prediction using synthetic minority oversampling technique and soft voting," in *2021 International Conference on Automation, Control and Mechatronics for Industry 4.0, ACMI 2021*, Jul. 2021, pp. 1–6, doi: 10.1109/ACMI53878.2021.9528188.

BIOGRAPHIES OF AUTHORS

Stuti Bhatt completed her diploma in information technology from G.P Srinagar Garhwal. She pursued her Bachelor's degree in IT from HNB Garhwal University followed by her post-graduation in computer science and engineering from Graphic Era Deemed to be University. Currently she an assistant professor at the Department of Computer Science and Engineering at Graphic Era Hill University, Dehradun, Uttarakhand, India. Her research interests span artificial intelligence, machine learning, 6G wireless communication networks, quantum computing, healthcare, blockchain, and cybersecurity. She can be contacted at email: stutibhatt1028@gmail.com.

Surender Reddy Salkuti received Ph.D. degree in electrical engineering from the Indian Institute of Technology (IIT), New Delhi, India, in 2013. He was a Postdoctoral Researcher at Howard University, Washington, DC, USA, from 2013 to 2014. He is currently an associate professor at the Department of Railroad and Electrical Engineering, Woosong University, Daejeon, Republic of Korea. His current research interests include market clearing, including renewable energy sources, demand response, and smart grid development with the integration of wind and solar photovoltaic energy sources. He can be contacted at email: surender@wsu.ac.kr.

Seong-Cheol Kim received B.S., M.S. and Ph.D. degrees, in electronic engineering from Korea University in 1987, 1989, and 1997, respectively. He is currently serving as Head of the Railroad and Electrical Engineering Department, Woosong University, Korea. His research interests are mobile communication systems and pulsed power systems. He can be contacted at email: kmin@wsu.ac.kr.