JRITM
JOURNAL OF RESEARCH AND INNOVATION IN TECHNOLOGY, COMMERCE AND MANAGEMENT

Journal of Research and Innovation in Technology, Commerce and Management
Vol. 3 Issue 9, September 2026, pp. 39118-39123
ISSN: 3049-3129 (Online)

A Review of Trust-Aware Explainable Federated Multimodal Learning for Early Cardiovascular Disease Prediction

Chandru K
PhD Research Scholar
Dept. of Information Technology
Government Arts College (Autonomous)
Coimbatore, Tamil Nadu, India
Email: chandrukannan419@gmail.com

D. Devakumari
Associate Professor
Dept. of Information Technology
Government Arts College (Autonomous)
Coimbatore, Tamil Nadu, India
Email: ddcbe04@gmail.com

Cite as: Chandru K, & D. Devakumari. (2026). A Review of Trust-Aware Explainable Federated Multimodal Learning for Early Cardiovascular Disease Prediction. Journal of Research and Innovation in Technology, Commerce and Management, Vol. 3(Issue 9), 39118–39123. https://doi.org/10.5281/zenodo.22868339

DOI: https://doi.org/10.5281/zenodo.22868339

Abstract

Cardiovascular disease (CVD) is the largest cause of death worldwide, and artificial intelligence can now detect it earlier than traditional methods. The difficulty is that the data needed to train a reliable model is held separately by individual hospitals and is protected by privacy law. Federated learning (FL) answers part of this problem by training a shared model without moving patient data out of each hospital. This review examines thirteen studies published between 2022 and 2026 and organises them around the four capabilities a usable CVD prediction system needs: multimodal data preparation, trust-aware aggregation, federated multimodal fusion, and explainable privacy-preserving decision support. The evidence is presented in table form for ease of comparison. Each capability has now been demonstrated on its own. Cross-attention fusion of ECG and clinical data outperforms ECG alone; reputation-weighted aggregation keeps a federated model accurate when some clients are unreliable; adaptive differential privacy reduces membership inference attack success to near chance; and SHAP and LIME explanations can now be scored for stability rather than simply displayed. No published framework, however,

combines all four and evaluates them on cardiovascular data under one protocol. This review sets out that gap and describes the four-phase framework intended to close it.

Keywords: Federated Learning; Cardiovascular Disease; Multimodal Fusion; Explainable Artificial Intelligence; Differential Privacy; Trust-Aware Aggregation; Electrocardiogram; Clinical Decision Support

1. Introduction

Cardiovascular disease accounts for roughly one third of deaths worldwide. Detecting it early matters because treatments such as lipid management, blood pressure control and lifestyle change work far better before permanent damage to the heart has occurred. The electrocardiogram (ECG) is cheap, quick and available almost everywhere, and deep learning models can read patterns in it that a human reader cannot see [3, 12]. Clinical records add a second and complementary source of information: age, cholesterol, blood pressure and diabetes status [1, 8].

The obstacle is data. A deep model needs records from many thousands of patients across different populations, and no single hospital holds enough of them. Privacy law prevents hospitals from pooling records into one central database. Federated learning was designed for this situation [7]. Each hospital trains the model on its own data, sends only the model updates to a central server, and the server combines those updates into a shared global model. Raw patient records never leave the hospital [3].

Federated learning alone, however, leaves three problems unsolved. First, hospital datasets are not statistically alike, a condition described as non-IID data, and standard averaging loses accuracy when this happens [10]. Second, standard averaging weights each hospital by how many records it holds, so a large but poorly labelled dataset counts for more than a small, carefully curated one, and a dishonest participant is not detected at all [2, 4]. Third, the resulting model is a black box, and clinicians will not act on a prediction they cannot interrogate [6, 13]. Alongside these problems runs an opportunity, because ECG data and clinical records describe different aspects of the same patient, so combining them predicts better than either one alone [1, 8].

This review therefore examines the evidence for a system that holds four capabilities at once. One is multimodal data preparation, meaning the cleaning, imputation and encoding needed to bring ECG images and clinical records into a single usable form. Second is trust-aware aggregation, meaning a rule that weights each hospital by how reliable its updates actually are. Third is federated multimodal fusion, meaning a model that learns from both modalities inside the federated setting and still works when one of them is missing. Finally, explainable and privacy-preserving decision support, meaning explanations a clinician can act on together with a formal privacy guarantee that has been tested against attack. The proposed study uses PTB-XL for ECG images and MIMIC-IV for clinical records, partitioned across simulated hospital clients, so the review pays particular attention to work using those or

comparable datasets [8, 9, 12].

2. Summary of the Reviewed Studies

Thirteen studies were selected for detailed review. Each study was then compared on its focus and method, its key result, and the limitation that leaves an opening for the present work. Table 1 sets out that comparison.

Table 1. *Comparison of the reviewed studies, 2022 to 2026*

#	Author Name, Year of publication and Journal Name	Focus and methods used	Key results	Limitations (gap for this study)
1	Stahlschmidt et al. (2022) [1], Briefings in Bioinformatics	Review of deep learning fusion strategies for biomedical data; proposes a taxonomy of early, intermediate (joint) and late fusion.	Deep fusion consistently beats single-modality and shallow models; joint representation learning is the preferred strategy.	A review only, not cardiovascular, and does not cover fusion inside a federated setting.
2	Rjoub et al. (2023) [2], Information Systems Frontiers	Trust-aware client selection for federated learning using deep reinforcement learning and transfer learning, tested on healthcare IoT devices.	Selecting clients on trust and resource cost gives a better accuracy-to-time trade-off than random selection.	Trust decides who takes part, not how much each client's update counts during aggregation. Not multimodal.
3	Weimann and Conrad (2024) [3], IEEE J. Biomedical and Health Informatics	Federated training of deep residual CNNs for ECG abnormality detection across several hospital databases (PhysioNet/CinC 2021).	Federated models beat hospitals training alone and nearly matched centralised training, including on unseen hospitals.	ECG only. Every hospital is trusted equally, no explanations, and no formal privacy guarantee.
4	Ghanta et al. (2025) [4], PeerJ Computer Science	FedAKCH: aggregation weights set by a reputation score from client performance, with CKKS homomorphic encryption of updates. Brain MRI.	99.39% accuracy, and 94% retained when half the clients were noisy, where standard FL fell to 33%.	Not cardiovascular, single modality, no explanations. Reputation uses past accuracy only.
5	Song et al. (2025) [5], Scientific Reports	Interpretable multimodal model combining ECG with acceleration signals, explained with SHAP.	The importance of ECG features rose with exercise intensity while accelerometer features dominated at moderate intensity.	Shows modality importance changes with context, so fixed fusion weights are wrong, but the model is centralised.
6	Kaj et al. (2025) [6], Concurrency and Computation: Practice and Experience	PKIFLIC-Net: multi-hospital decision support with FedNova attention weighting, differential privacy, homomorphic encryption and SHAP. MIMIC-IV.	88.4% accuracy and AUC 0.93, with SHAP explanations that clinicians could review and override.	Electronic records only, no ECG. Attention weights are not a formal trust score and explanations are not pooled across hospitals.
7	Kyu et al. (2025) [7], Healthcare	Comparative review of 28 federated learning studies for CVD, split into biosignal-based and EHR-based approaches.	Biosignal work suits personalised monitoring; EHR work suits large hospital consortia. The two are complementary.	Very few studies combine the two data types, and none combine them with trust scoring and explanation.
8	Qu et al. (2026) [8], PLOS ONE	MAF-Net: multi-scale CNN and bi-LSTM for ECG, a feature-crossing branch for clinical data, joined by two-way cross-attention. PTB-XL.	Top-line score of 90.75% and macro AUC 0.9407 across five cardiac classes, above ECG-only baselines.	All data pooled centrally, so there is no privacy protection and no multi-hospital use.
9	Liu et al. (2026) [9], J. Imaging Informatics in Medicine	FedFed-Cardio: semi-federated CVD risk stratification over cardiac imaging, physiological signals and EHR, explained with LIME and Grad-CAM.	AUC 0.972 across five simulated institutions, 28% less communication and F1 above 92% under noise.	Only 125 patients. All clients trusted equally, explanations stay local, and no differential privacy or attack testing.
10	Sengan and Shieh (2026) [10], Scientific Reports	Non-IID-aware federated framework: client clustering by distribution similarity, gradient correction, geometric-median filtering, reputation control.	Kept 85.61% accuracy with four dishonest clients, and handled quantity, label and feature skew at the same time.	Not cardiovascular and not multimodal. Trust is several separate defences rather than one score.
11	Adhi Siva and Chowdhary (2026) [11], Scientific Reports	Personalised multimodal federated learning with Vision Transformers, KL-divergence personalisation masks and adaptive differential privacy.	Converged in 47 rounds against 85 for FedAvg, and cut membership inference attack success to 52.1% (chance = 50%).	Imaging only, so no ECG-plus-records fusion, and client reliability is never scored.
12	Czerwinski et al. (2026) [12], Computer Methods and Programs in Biomedicine	CVD detection from ECG printout images with heavy augmentation, plus LIME explanations scored for stability by Intersection over Union.	Training on deliberately damaged images gave AUC 0.894 on a contaminated test set and the most stable explanations (IoU 0.399).	Single site, single modality, no federation and no privacy mechanism.
13	Houpeikis et al. (2026) [13], IEEE Trans. Biomedical Engineering	Post-hoc explanation of ECG CNNs by clustering feature activations into time-localised segments.	Recovered P, QRS, T and IP landmarks; a surrogate model on cluster proportions reproduced 94.9% of CNN predictions.	Centralised and single modality. Explanations are computed on one dataset, not reconciled across hospitals.

3. Discussion

3.1 Multimodal data preparation

Preparation decides what the model can learn. Stahlschmidt et al. [1] show across the biomedical literature that fusing modalities beats using one, and that joint representation learning is usually the strongest option. Qu et al. [8] confirm this specifically for CVD: their cross-attention model, which lets clinical variables guide which parts of the ECG the network attends to, reached 90.75% on PTB-XL, above ECG-only baselines. Song et al. [5] add an important qualification. Using SHAP, they found that the importance of ECG features rose with exercise intensity while other signals dominated at moderate intensity, which means modality importance changes with context and fixed fusion weights are inappropriate. Czerwinski et al. [12] cover a practical detail that is easy to overlook: ECGs in real hospitals are often stored as printed or scanned images carrying creases, handwriting and discolouration, and models trained only on clean images fail on them. Training on deliberately damaged images raised performance on a contaminated test set to AUC 0.894.

3.2 Trust-aware aggregation

Three studies show that treating every hospital as equally reliable is a design flaw, and each solves it differently. Rjoub et al. [2] use trust to decide which clients are invited into a training round at all. Ghanta et al. [4] go further and set the aggregation weight itself from a reputation score, retaining 94% accuracy with half the clients noisy where standard federated averaging collapsed to 33%. Sengan and Shieh [10] assemble a layered defence of distribution-based clustering,

gradient correction, geometric-median filtering and reputation control, keeping 83.61% accuracy with four dishonest clients. What none of them provides is a single, formally defined trust score that measures distributional deviation, update alignment and predictive competence together, and none has been tested on cardiovascular multimodal data.

3.3 Federated multimodal fusion

Weimann and Conrad [3] establish the baseline: federated ECG classifiers beat hospitals training alone and come close to centralised training, including on hospitals that never joined. Their model uses one modality. Liu et al. [9] come closest to the target architecture, fusing imaging, physiological signals and records across five simulated institutions with an AUC of 0.972, but on only 123 patients and with every client trusted equally. Adhi Siva and Chowdhary [11] personalise the federated model using KL-divergence masks and converge in 47 rounds against 85 for FedAvg, though they work on images rather than ECG and records. A complication none of them addresses is that modality availability is itself uneven: one hospital may hold ECGs and full laboratory panels, another ECGs only. A fusion model that assumes both modalities are present everywhere will fail at exactly the hospitals that matter most for generalisability, which is what modality dropout during training is intended to prevent.

3.4 Explainability and privacy

Raj et al. [6] come closest to a complete decision support system, running across hospitals on MIMIC-IV with differential privacy, homomorphic encryption and SHAP explanations clinicians can respond to, at 88.4% accuracy and AUC 0.93. It uses records only. Adhi Siva and Chowdhary [11] contribute the stricter privacy standard: instead of reporting only a privacy budget, they tested whether an attacker could tell if a specific patient was in the training data and reduced that success rate to 52.1% against a chance level of 50%. Czerwinski et al. [12] treat explanation stability as a measurable quantity, scoring how consistently LIME highlights the same region when the input is perturbed, and Boubekki et al. [13] make explanations anatomically meaningful by recovering P, QRS, T and TP landmarks from the network's own features. Explaining a federated model raises a problem with no centralised equivalent: local explanations are computed on local data, so a feature that dominates at one hospital may be irrelevant at another, and averaging SHAP values without accounting for this describes no real patient population.

4. Research Gap

Each of the four capabilities has been demonstrated, but always separately, and usually on a different disease, modality or dataset. Trust-aware aggregation exists but not for cardiovascular multimodal data [2, 4, 10]. Cross-attention fusion of ECG and clinical data exists but only when the data is centralised [8]. Federated explanation aggregation is named as a requirement but has no published formal treatment with cardiovascular validation [6, 7, 9]. Attack-tested differential privacy exists but has not been combined with either trust scoring or multimodal fusion [11]. Table 2 maps the research gaps

Table 2. *Research Gaps*

Citations	What is still missing	Contribution of this study
Stahlschmidt et al. [1]; Song et al. [5]; Qu et al. [8]; Czerwinski et al. [12]	ECG images and clinical records are prepared separately, and no study splits a combined dataset into realistic non-IID hospital partitions.	One pipeline that cleans, imputes and encodes PTB-XL ECG images together with MIMIC-IV clinical records, then partitions them across simulated hospitals.
Rjoub et al. [2]; Ghanta et al. [4]; Sengan and Shieh [10]	Trust either selects clients or uses past accuracy alone, and has never been tested on cardiovascular multimodal data.	One trust score per round combining KL-divergence, gradient cosine similarity and validation loss, used directly as the aggregation weight.
Weimann and Conrad [3]; Qu et al. [8]; Liu et al. [9]; Adhi Siva and Chowdhary [11]	Cross-attention fusion works only when data is centralised, and federated studies stay single-modality or image-only.	ResNet-18 for ECG images and an MLP for clinical features joined by cross-modal attention, trained with modality dropout so missing data is tolerated.
Raj et al. [6]; Adhi Siva and Chowdhary [11]; Czerwinski et al. [12]; Boubekki et al. [13]	SHAP and LIME are computed locally at each hospital, and privacy is usually reported as an epsilon value without attack testing.	Federated SHAP aggregation with LIME, DP-SGD at epsilon = 1, 3 and 5, validated by membership inference attack testing and explanation stability analysis.

5. Conclusion

Federated learning is now the accepted way to train cardiovascular prediction models across hospitals, and the studies reviewed here show that its known weaknesses can each be addressed. Reputation-weighted and distribution-aware aggregation survives unreliable participants and unequal data. Cross-attention fusion of ECG and clinical data outperforms single-modality prediction. Adaptive differential privacy can defeat membership inference attacks without destroying accuracy. Explanations can be scored for stability and anchored in cardiac anatomy. What has not been done is to combine these four capabilities in one system and evaluate it on cardiovascular data under a single protocol, which is the contribution the proposed trust-aware explainable federated multimodal framework is intended to make. How it is evaluated matters just as much, because the reviewed literature relies heavily on retrospective data and simulated hospital splits, and a more sophisticated system that is not evaluated more rigorously would have no stronger claim to clinical trust.

References

1. Stahlschmidt SR, Ulfenborg B, Synnergren J. Multimodal deep learning for biomedical data fusion: a review. Briefings in Bioinformatics. 2022;23(2):bbab569. https://doi.org/10.1093/bib/bbab569

2. Rjoub G, Wahab OA, Bentahar J, Cohen R, Bataineh AS. Trust-augmented deep reinforcement learning for federated learning client selection. Information Systems Frontiers. 2022 (published 2024;26(4):1261-1278). https://doi.org/10.1007/s10796-022-10307-z

3. Weimann K, Conrad TOF. Federated learning with deep neural networks: a privacy-preserving approach to enhanced ECG classification. IEEE Journal of Biomedical and Health Informatics. 2024;28(11). https://doi.org/10.1109/JBHI.2024.3427787

4. Ghanta S, Boyapati P, Biswas S, Pradhan AK, Mohanty SP. Enhancing privacy-preserving brain tumor classification with adaptive reputation-aware federated learning and homomorphic encryption. PeerJ Computer Science. 2025;11:e3165. https://doi.org/10.7717/peerj-cs.3165

5. Song Y, Wang Z, Wang H, Sun G, Gong B, Zhang F. Interpretable deep learning for personalised energy expenditure prediction using ECG and acceleration signals in incremental exercise. Scientific Reports. 2025;15(1):36277. https://doi.org/10.1038/s41598-025-20195-3

6. Raj G, et al. A privacy-preserving and interpretable federated learning framework with clinician feedback for multi-hospital clinical decision support. Concurrency and Computation: Practice and Experience. 2025;37(27-28):e70365. https://doi.org/10.1002/cpe.70365

7. Ryu H, Lee M, Kim S-H, Kim JH, Yang H-J. Federated learning for cardiovascular disease prediction: a comparative review of biosignal- and EHR-based approaches. Healthcare. 2025;13(21):2811. https://doi.org/10.3390/healthcare13212811

8. Qu C, Zhang X, Lu Y, Wang Y, Su C. MAF-Net: multimodal cross-attention-based fusion network for cardiovascular disease classification. PLOS ONE. 2026;21(4):e0345238. https://doi.org/10.1371/journal.pone.0345238

9. Liu X, Li S, Zhu Q, et al. Interpretable semi-federated learning for multimodal cardiac imaging and risk stratification: a privacy-preserving framework. Journal of Imaging Informatics in Medicine. 2026;39:2641-2660. https://doi.org/10.1007/s10278-025-01643-y

10. Sengan S, Shieh C-S. Non-IID and aware federated intrusion detection with PBFT with secured model aggregation for multi-institutional healthcare internet of things networks. Scientific Reports. 2026;16(1). https://doi.org/10.1038/s41598-026-56167-4

11. Adhi Siva M, Chowdhary CL. Multi-modal personalised federated learning with adaptive differential privacy for medical image classification and a privacy-preserving approach. Scientific Reports. 2026;16(1). https://doi.org/10.1038/s41598-026-49896-z

12. Czerwinski A, Kucharski D, Kawa J, Zheng Y, Lip GYH, Nalepa J, Wijata AM. Detecting cardiovascular diseases using ECG scans and explainable artificial intelligence. Computer Methods and Programs in Biomedicine. 2026;283:109428. https://doi.org/10.1016/j.cmpb.2026.109428

13. Boubekki A, Patlatzoglou K, Barker J, Aggour H, Ng FS, Ribeiro AH. Explaining deep learning for ECG using time-localised clusters. IEEE Transactions on Biomedical Engineering. 2026. https://doi.org/10.1109/TBME.2026.3702945