International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0471
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2971 - 2977

A Comprehensive Review of Federated Learning and Explainable AI Approaches for Privacy-Preserving Breast Cancer Detection: Advancements in Multi-Modal Data Fusion, Interpretability, and Clinical Trust

Sandhya H[1], Dr. Kirubha D[2], Dr. T.C.Manjunath[3]
[1]PG MTech. Student, USN - 1RR23SCS01, Second Year, Dept. of Computer Science & Engineering, Rajarajeswari College of Engineering, Bangalore, Karnataka, India.
[2]Project Guide, Professor and HoD, Dept. of Computer Science & Engineering, Rajarajeswari College of Engineering, Bangalore, Karnataka, India.
[3]Dean Research (R & D), Professor, Dept. of Computer Science & Engineering, IoT Cyber Security & Blockchain Technology, Rajarajeswari College of Engineering, Bangalore, Karnataka, India.
Email: *tcmanju@iitbombay.org[3]*

Abstract
In this review paper, a comprehensive review of federated learning and explainable ai approaches for privacy-preserving breast cancer detection is presented. Advancements in Multi-Modal Data Fusion, Interpretability, and Clinical Trust. Breast cancer remains one of the leading causes of mortality among women worldwide, demanding diagnostic solutions that are accurate, secure, and clinically interpretable. With the rapid growth of artificial intelligence in healthcare, federated learning (FL) and explainable AI (XAI) have emerged as complementary paradigms that address two critical aspects: preserving patient privacy while ensuring transparency in decision-making. This review paper explores how FL enables collaborative model training across distributed clinical institutions without centralizing sensitive medical data, while XAI frameworks enhance trust by making predictions understandable for clinicians. A special emphasis is placed on multi-modal data fusion—integrating mammography, histopathology, genomics, and electronic health records—to improve diagnostic robustness and reliability. The paper synthesizes current advancements, challenges, and future directions, highlighting how privacy-preserving FL techniques, combined with interpretable models, can foster clinical trust and adoption in real-world healthcare systems. By comparing state-of-the-art approaches, the review provides insights into algorithmic performance, interpretability trade-offs, and ethical considerations in medical AI. Ultimately, this work underscores the importance of balancing technological innovation with clinical usability, positioning federated learning and explainable AI as pivotal enablers for the next generation of breast cancer detection systems.

Keywords: *Federated learning, Explainable AI, Privacy-preserving healthcare, Breast cancer detection, Multi-modal data fusion, Interpretability, Clinical trust, Medical AI, Healthcare ethics, Secure machine learning.*

1. Introduction
Breast cancer continues to be one of the most pressing health challenges worldwide, with early and accurate detection playing a vital role in improving patient survival and treatment outcomes. The rapid growth of artificial intelligence (AI) has opened new avenues for enhancing diagnostic accuracy, particularly using machine learning and deep learning models that can analyze complex medical data. However, two critical barrier-patient data privacy and model transparency—have limited the widespread clinical adoption of AI-based diagnostic tools. Federated Learning (FL) has emerged as a promising solution to the privacy dilemma by enabling collaborative model training across distributed

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0471
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2971 - 2977

institutions without requiring direct data sharing. At the same time, Explainable AI (XAI) addresses the "black box" nature of AI models by making their predictions more transparent and interpretable to clinicians. Together, these approaches hold the potential to deliver AI systems that are not only accurate but also trustworthy and ethically sound. Furthermore, the integration of multi-modal data fusion—combining imaging, histopathology, genomics, and electronic health records—offers a more holistic view of breast cancer diagnosis and prognosis, strengthening both predictive performance and clinical confidence. This review paper provides a comprehensive synthesis of current advancements in federated learning and explainable AI applied to breast cancer detection. It highlights the latest research trends, methodologies, and challenges, while emphasizing the importance of interpretability, privacy preservation, and clinical trust in shaping the future of AI-driven healthcare solutions [11-15].

2. Works Done by Various Authors

Breast cancer detection has increasingly benefited from the fusion of advanced machine learning models and explainable AI (XAI) techniques, both of which aim to enhance diagnostic accuracy while ensuring transparency. Gupta and Sharma focused on addressing class imbalance using SMOTE and Tomek Links, evaluating boosting algorithms such as Light GBM, Cat Boost, and XGBoost on the SEER dataset. Their findings revealed Light GBM as the most effective, with LIME explanations adding interpretability and strengthening clinical trust in predictions. Complementing this, Ardabil et al. reviewed a large body of deep learning studies, observing the dominance of CNNs in imaging tasks, as well as a growing trend toward transfer learning and ensemble methods. They emphasized the role of multi-modal data integration and XAI in achieving reliable, clinician-friendly diagnostic tools. In a similar line, Tanim et al. designed a lightweight neural network by blending deep learning and traditional machine learning, achieving an impressive accuracy of 97.54%. They used SHAP and LIME to interpret outputs, making the model suitable for real-time clinical application. Likewise, Maheswari et al. found that Random Forest, combined with PCA and SMOTE preprocessing, outperformed other tested models with 95.9% accuracy. The addition of SHAP and LIME explanations further improved interpretability and clinician acceptance. Extending this to ultrasound imaging, Butt and Asif proposed an ensemble of Mobile Net and Xception, supported by Grad-CAM visualizations, achieving an AUC close to 0.98—demonstrating the power of combining lightweight architectures with XAI-driven visual insights. Ariyametkul et al. reinforced this trend by testing six CNNs, with DenseNet201 achieving approximately 99% accuracy. They employed LIME and Grad-CAM to enhance transparency, which strengthened the model's clinical applicability. Jain et al. contributed through Auto ML with TPOT and PCA on the Wisconsin dataset, achieving 98.6% accuracy. While their pipeline lacked direct XAI integration, they pointed to the importance of embedding SHAP or LIME in future iterations for better clinical acceptance. Similarly, Basha et al. found Random Forest to be most effective among traditional algorithms, but improved overall accuracy through stacked ensemble learning, interpreting results with SHAP to highlight feature importance. Beyond classification, survival prediction has also seen advancements. Hashtarkhani et al. developed a multi-level survival model using EMR data fused with socioeconomic and environmental factors, with SHAP offering valuable explanations for disparities in outcomes, paving the way for fairer healthcare decisions. Finally, Zhang et al. introduced XAI-RACaps Net, an innovative model combining Transformers with Capsule Networks and relevance heat maps. Their approach outperformed conventional CNNs while providing clinically meaningful explanations, marking a step forward in explainable and high-performance diagnostic models. Table 1 gives the comparisons. Together, these studies highlight the strong synergy between high-performing algorithms and XAI frameworks. While accuracy remains a primary focus, the integration of interpretability tools such as SHAP, LIME, and Grad-CAM is becoming indispensable for building trust and ensuring that AI solutions in breast cancer detection are not only powerful but also transparent and clinically reliable [16-21].

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0471
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2971 - 2977

Table 1 Comparison of the Works Done by Authors

Author	Method / Algorithm	Dataset	Accuracy / AUC	XAI Technique	Key Contribution
Gupta & Sharma [1]	LightGBM, CatBoost, XGBoost + SMOTE + Tomek Links	SEER dataset	Best: LightGBM (High recall & accuracy)	LIME	Tackled class imbalance; boosted transparency in predictions
Ardabili et al. [2]	CNNs, Transfer Learning, Ensemble Models	Multiple imaging studies	–	– (Recommended XAI use)	Comprehensive review of DL in breast cancer; stressed multi-modal + XAI integration
Tanim et al. [3]	Lightweight Neural Network + Hybrid DL & ML	–	97.54%	SHAP, LIME	Achieved high accuracy, real-time applicability, outperformed traditional classifiers
Maheswari et al. [4]	Random Forest with PCA + SMOTE preprocessing	–	95.9%	LIME, SHAP	RF outperformed other ML models; enhanced interpretability for clinicians
Butt & Asif [5]	MobileNet + Xception Ensemble	Ultrasound images	AUC ~0.98	Grad-CAM	Improved ultrasound tumour detection with strong visualization support
Ariyametkul et al. [6]	DenseNet201 among 6 CNNs	–	~99%	LIME, Grad-CAM	Highest performance CNN; interpretability tools increased clinical trust
Jain et al. [7]	AutoML (TPOT) + PCA	Wisconsin dataset	98.6%	– (Planned SHAP/LIME)	Automated pipeline optimization; scope for future XAI integration
Basha et al. [8]	Random Forest + Stacked Ensemble	–	High accuracy	SHAP	Highlighted feature importance; improved ensemble accuracy
Hashtarkhani et al. [9]	Survival prediction with EMR + socioeconomic & environmental data	EMR datasets	– (Survival outcomes)	SHAP	Identified disparities; explained feature impact on survival outcomes
Zhang et al. [10]	XAI-RACapsNet (Capsule Net + Transformers + Relevance Heat Maps)	–	Outperformed CNNs	Relevance Heat Maps	Advanced capsule-based model; improved clinical insights with strong explanations

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0471
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2971 - 2977

3. Survey Extended

Breast-cancer FL has moved from prototypes to increasingly realistic multi-site evaluations. Almufareh et al. proposed an end-to-end FL pipeline (detection → localization → classification) tailored for breast cancer images, showing that decentralized training can deliver competitive accuracy without pooling data. Building on the privacy side, Shukla et al. fused FL with differential privacy to curb membership-inference risks while maintaining diagnostic performance, and Alhaji et al. introduced an explainable federated vision-transformer framework, coupling global modeling with local interpretability for breast prediction tasks. Complementing single-paper advances, the ACR–NCI–NVIDIA federation challenge analysis by Schmidt et al. reported that FL models for breast density generalize better than single-site learners—a recurring motivation for cross-institutional training. Deployment isn't just an algorithmic issue; it's operational. Brink et al. documented the nuts and bolts of multi-institutional development for a breast-density algorithm—IRB timelines, data heterogeneity, and site-specific quirks—offering a realistic view of what it takes to translate from lab to clinic across hospitals. Beyond FL, privacy-preserving inference is gaining traction for longitudinal and survivorship tasks. Son et al. demonstrated homomorphic encryption + secure two-party computation for a GRU-based breast-cancer recurrence predictor—proof that sensitive temporal models can be run without exposing raw inputs. More recently, Selvakumar et al. used TFHE with quantization-aware training to make encrypted inference feasible, sketching a path to compliant cloud workflows for oncology AI. Explainability in mammography continues to mature from heat-maps to measurable, annotation-aware explanations. Talaat et al. integrated Grad-CAM into an Inception-ResNet v2 pipeline for mammograms, reporting clearer saliency around suspicious regions; Camurdan et al. tackled the labeling bottleneck with a patch-based curriculum and explicit XAI checks under limited strong labels; and Sreekala et al. combined Grad-CAM, SHAP, LIME, and LRP in one framework, arguing that ensembles of explainers give radiologists richer, cross-validated evidence. Several teams have stressed that XAI must be quantified, not just visualized. Ahmed et al. (arXiv) coupled CNNs with XAI and used the Hausdorff distance to gauge how well explanations align with expert annotations—an evaluation step that's often missing. Meanwhile, Ghasemi et al.'s scoping review found SHAP to be the most widely adopted model-agnostic technique across breast-cancer tasks, especially for tree/ensemble models. Methodological reviews by Shifa et al. and Ansari et al. mapped how explanation quality, clinical interpretability, and decision support together, with concrete guidance for imaging pipelines [22-26].

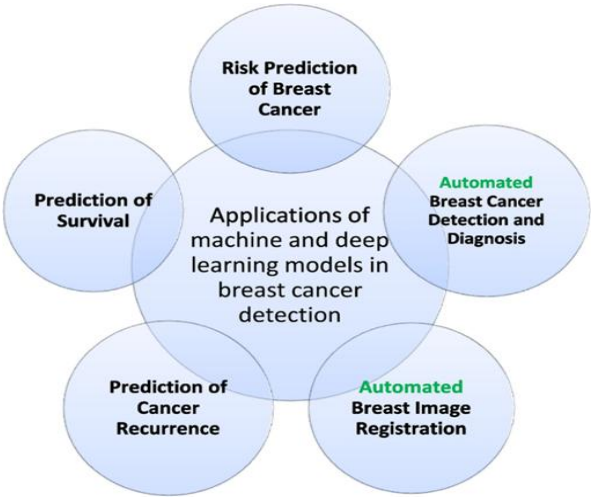

Figure 1 A Conceptual View of the Breast Disease overview [1]

Figure 1 gives a conceptual view of the breast disease overview [1]. Ultrasound has seen parallel momentum. Latha et al. fine-tuned EfficientNet variants and used Grad-CAM to make lesion cues explicit; Oztekin et al. compared explainable pipelines for benign vs. malignant solid masses; and Moldovanu et al. combined LIME + SHAP to interpret lesion classification, highlighting how different explainers surface complementary error modes in breast ultrasound. Histopathology and subtyping continue to benefit from XAI-aware modeling. Thatha et al. proposed a DL system for subtype classification with interpretability hooks for early triage; Ukwuoma et al. built a histopathological classifier spanning breast/colon/lung images and

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0471
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2971 - 2977

emphasized XAI for pathologist validation; and Luong et al.'s PCA-assisted ViTs incorporated SHAP/LIME/Grad-CAM to pinpoint discriminative tissue structures. Multi-modal fusion is increasingly central to real-world decision support. Rabah et al. integrated mammography with clinical metadata to classify lesions into five categories, an example of pragmatic fusion that matches the information mix radiologists use. On the review side, Abdullakutty et al. surveyed histopathology–genomics–EHR combinations and detailed early/late/attention-based fusion designs; Al-Tam et al. built a hybrid explainable CAD pipeline with Grad-CAM to reconcile image- and metadata-driven cues at detection and refinement stages. Generalizable lessons about clinical trust are emerging from user studies. Rezaeian et al. found that more explanation isn't always better—trust and accuracy can plateau or even dip with dense rationales—while a broader experiment on clinical decision-support systems (CDSS) showed that the level and form of explanations shape both trust and diagnostic performance in breast workflows. A parallel HCI-driven study reported nuanced trade-offs among trust, cognitive load, and performance, underscoring the need for human-factors evaluation alongside ROC/AUC. Finally, hybrid and integrative models continue to push accuracy while keeping the "why" visible. Alom et al. demonstrated an integrative hybrid DL approach spanning CBIS-DDSM and WBCD with Grad-CAM overlays for radiologist cross-checks; Sreekala et al. and Ahmed et al. (above) converge on the idea that multi-explainer or metric-validated XAI improves clinician confidence. At the molecular level, Chereda et al.'s Graph-CNN + GLRP method generated patient-specific subnetworks for metastasis prediction—an example of explainability for omics that can dovetail with imaging in multi-modal FL settings. Lamy et al.'s visual case-based reasoning for therapy planning remains a touchstone for case-level interpretability in breast cancer [27-30].

Conclusions

Across imaging, pathology, and molecular data, the field is coalescing around three pillars: (1) federation (to share signal, not data), (2) privacy-preserving inference (DP/FHE), and (3) auditable explanations (Grad-CAM/SHAP/LIME + objective quality metrics). The most promising systems combine these pillars with multi-modal fusion and prospective human-factors testing, which is exactly the space your review targets. This review highlights that federated learning (FL) and explainable AI (XAI) together form a strong foundation for advancing privacy-preserving breast cancer detection. FL enables collaborative model training across institutions without exposing sensitive medical records, addressing one of the key barriers to clinical deployment. At the same time, XAI approaches such as SHAP, LIME, and Grad-CAM improve transparency and interpretability, allowing clinicians to understand and trust model outputs rather than relying on opaque "black-box" predictions. Across the literature, boosting algorithms, CNNs, ensemble models, capsule networks, and transformer-based architectures have all been enriched by XAI techniques to deliver both high diagnostic accuracy and clinical usability. Multi-modal data fusion—integrating imaging, pathology, genomics, and electronic health records—has further shown promise in strengthening prediction robustness and patient-specific insights. Importantly, studies also caution that explanation quality and clinician trust must be carefully balanced, with evidence that too much or poorly designed explanation can diminish user confidence. Overall, the synthesis of evidence suggests that the most impactful solutions for breast cancer detection will come from systems that combine federated training for privacy, explainable pipelines for trust, and multi-modal integration for robustness. Future research should focus on large-scale federated trials, standardized evaluation of explanation quality, and human-factors studies that assess how clinicians interact with AI explanations in real decision-making contexts. These directions will be critical for moving from promising research to safe, ethical, and widely accepted deployment of AI in clinical breast cancer care.

References

[1]. Thakur, Neha & Kumar, Pardeep & Kumar, Amit. (2023). A systematic review of machine and deep learning techniques for the

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0471
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2971 - 2977

identification and classification of breast cancer through medical image modalities. Multimedia Tools and Applications. 83. 1-94. 10.1007/s11042-023-16634-w.

[2]. Vinayak Gupta, Richa Sharma, "Enhancing Breast Cancer Prediction with XAI Enabled Boosting Algorithms".

[3]. Advances of Deep learning in Breast Cancer Modeling Sina Ardabili, Ardabil, Amirhosein Mosavi, Imre Felde

[4]. Breast Cancer Diagnosis with XAI-Integrated Deep Learning Approach, Sharia Arfin Tanim, Tahmid Enam Shrestha, Fariha Jahan

[5]. Interpretable Machine Learning Model for Breast Cancer Prediction Using LIME and SHA, P B Uma Maheswari, Aaditi A, Ananya Avvaru, Aryan Tandon, R. Pérez de Prado

[6]. Beyond Boundaries: A Novel Ensemble Approach for Breast Cancer Detection in Ultrasound Imaging Using Deep Learning, Ateeq Ur Rehman Butt, Muhammad Asif

[7]. Explainable AI (XAI) for Breast Cancer Diagnosis, Awika Ariyametkul; Sudarshan Tamang; May Phu Paing

[8]. Advanced Machine Learning Techniques for Breast Cancer Detection and Classification with Explainable AI, Tanveer Basha; Mohammed Ziyanuddin; Niranjana Sampathila; Hilda Mayrose; Krishnaraj Chadaga

[9]. AI4BCancer: Breast Cancer Classification using AutoML - TPOT with PCA, Charvi Jain; Shaurya Singh Srinet; Tarush Kumar Goyal; Karpagam M

[10]. An Explainable AI Data Pipeline for Multi-Level Survival Prediction of Breast Cancer Patients Using Electronic Medical Records and Social Determinants of Health Data, Soheil Hashtarkhani; Shelley White-Means; Sam Li; Rezaur Rashid; Fekede Kumsa; Cindy Lemon

[11]. BI-RADS-NET-V2: A Composite Multi-Task Neural Network for Computer-Aided Diagnosis of Breast Cancer in Ultrasound Images with Semantic and Quantitative Explanations Boyu Zhang, Aleksandar Vakanski, Andminxian

[12]. Breast Cancer Diagnosis: A Comprehensive Exploration of Explainable Artificial Intelligence (XAI) Techniques Samita Bai1, Sidra Nasir1, Rizwan Ahmed Khan, Sheeraz Arif1, Alexandre Meyer, and Hubert Konik

[13]. A novel approach for breast cancer detection using optimized ensemble learning framework and XAI, Raafat M. Munshi, Muhammad Umer, Lucia Cascone, Nazik Alturk, Oumaima Saidani, Amal Alshardan

[14]. XAI-driven CatBoost multi-layer perceptron neural network for analyzing breast cancer P. Naga Srinivasu, G. Jaya Lakshmi, Abhishek Gudipalli, Sujatha Canavoy Narahari, Jana Shafi, Marcin Woźniak & Muhammad Fazal Ijaz

[15]. Enhancing Breast Cancer Diagnosis in Mammography:Evaluation and Integration of Convolutional Neural Networks and Explainable AI, Maryam Ahmed, Tooba Bibi, Rizwan Ahmed Khan, and Sidra Nasir

[16]. Enhancing Breast Cancer Diagnosis in Mammography: Evaluation and Integration of Convolutional Neural Networks and Explainable AI Maryam Ahmed, Tooba Bibi, Rizwan Ahmed Khan, and Sidra Nasir

[17]. XAI-RACapsNet: Relevance aware capsule network-based breast cancer detection using mammography images via explainability O-net ROI segmentation Ahmed Alhussen Seifedine Kadry, Mohd Anul Haq, Arfat Ahmad Khan, Rakesh Kumar Mahendran

[18]. Ghasemi, Arash Shaban- Soheil Hashtarkhani, David L. Schwartz Explainable artificial intelligence in breast cancer detection and risk prediction: A systematic scoping review Amirehsan

[19]. Advancing Ovarian Cancer Diagnosis Through Deep Learning and eXplainable AI: A Multiclassification Approach Meera Radhakrishnan H. Muralikrishna, Niranjana Sampathila, (Senior Member, Ieee), Andk.S. Swathi

[20]. M. F. Almufareh, N. Tariq, M. Humayun, and B. Almas, "A Federated Learning Approach

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0471
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2971 - 2977

to Breast Cancer Prediction in a Collaborative Learning Framework," Healthcare, vol. 11, no. 24, Art. 3185, 2023.

[21]. A. R. Dandekar, A. Sharma, and J. K. Mishra, "Optimized Federated Learning Algorithm for Breast Cancer Detection Using the Marine Predators Algorithm," Journal of Electrical Systems, vol. 20, no. 1 s, pp. 911–920, 2024.

[22]. E. H. Houssein, M. M. Emam, and A. A. Ali, "An optimized deep learning architecture for breast cancer diagnosis based on improved marine predators algorithm," Neural Computing and Applications, vol. 34, no. 20, pp. 18015–18033, 2022.

[23]. M. Al-Hejri, A. H. Sable, R. M. Al-Tam, M. A. Al-Antari, S. S. Alshamrani, K. M. Alshmrany et al., "A hybrid explainable federated-based vision transformer framework for breast cancer prediction via risk factors," Scientific Reports, vol. 15, Art. 18453, 2025.

[24]. M. F. Almufareh, N. Tariq, M. Humayun, and B. Almas, "A Federated Learning Approach to Breast Cancer Prediction in a Collaborative Learning Framework," Healthcare (Basel), vol. 11, no. 24, Art. 3185, Dec. 2023, doi: 10.3390/healthcare11243185.

[25]. A. M. Al-Hejri, A. H. Sable, R. M. Al-Tam, M. A. Al-Antari, S. S. Alshamrani, K. M. Alshmrany, and W. Alatebi, "A Hybrid Explainable Federated-Based Vision Transformer Framework for Breast Cancer Prediction via Risk Factors," Scientific Reports, vol. 15, Art. 18453, May 2025, doi: 10.1038/s41598-025-96527-0.

[26]. J. Peta and S. Koppu, “Enhancing Breast Cancer Classification in Histopathological Images Through Federated Learning Framework,” IEEE Access, vol. 11, pp. 61866–61880, 2023,

[27]. D. Truhn et al., “Encrypted Federated Learning for Secure Decentralized Collaboration in Cancer Image Analysis,” Medical Image Analysis, vol. 92, art. no. 103059, 2024, doi: 10.1016/j.media.2023.103059.

[28]. A. Bechar, Y. Elmir, Y. Himeur, R. Medjoudj, and A. Amira, “Federated and Transfer Learning for Cancer Detection Based on Image Analysis,” Neural Computing and Applications, vol. 37, pp. 2239–2284, 2025.

[29]. F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “A Dataset for Breast Cancer Histopathological Image Classification,” IEEE Transactions on Biomedical Engineering, vol. 63, pp. 1455–1462, 2015.

[30]. A. M. Al-Hejri et al., “A Hybrid Explainable Federated-Based Vision Transformer Framework for Breast Cancer Prediction via Risk Factors,” Scientific Reports, vol. 15, art. no. 18453, May 2025