International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhancing Breast Cancer Detection and Prognosis through AI/ML-Based Algorithms

Sesha Bhargavi Velagaleti[1], Reshma Rege[2], Wasim Ahmad Bhat[3], Mehreen Shehzadi Khan[4], Dr. Sachin Malhotra[5], Dheeraj Kumar Sahni[6], P. Varaprasada Rao [7]

Submitted: 23/11/2023 **Revised**: 27/12/2023 **Accepted**: 08/01/2024

Abstract. Breast cancer, particularly when it has spread to other regions of the body, presents substantial treatment and prognosis concerns. Researchers have been in the forefront of developing Artificial Intelligence/Machine Learning (AI/ML)-based systems to solve these difficulties. When compared to traditional approaches, these new algorithms provide a viable route for identifying breast cancer with more accuracy and efficiency. In this study, we look at the creation and evaluation of AI/ML-based algorithms for improving breast cancer detection and prognosis. We investigate how these algorithms use cutting-edge technology to increase breast cancer diagnostic accuracy, especially in complicated and advanced stages of the illness. Additionally, we investigate how these algorithms contribute to a better understanding of the prognosis for breast cancer patients, enabling more tailored treatment plans. Our study demonstrates the potential of AI/ML-driven solutions to revolutionize breast cancer detection and prognosis. Through the incorporation of large datasets, advanced image analysis techniques, and predictive modeling, these algorithms offer a significant advancement in the field of oncology. We present evidence of their efficacy, highlighting the crucial role they play in early diagnosis, more accurate prognosis, and ultimately, improved patient outcomes. This research serves as a valuable contribution to the ongoing efforts to combat breast cancer and underscores the transformative potential of AI/ML-based algorithms in the realm of healthcare and disease management.

Keywords: *Breast cancer, AI/ML-based algorithms, Accuracy, Efficiency, Predictive modeling, Healthcare*

1 Introduction

Breast cancer, as one of the most prevalent and devastating forms of cancer, continues to present formidable challenges in the realm of healthcare. Its complexity is further amplified when it progresses to metastatic stages, spreading to other parts of the body. Metastatic breast cancer not only necessitates a more aggressive and multifaceted treatment approach but also significantly impacts the prognosis, often leading to a more challenging clinical course. In this context, the development of innovative and transformative solutions is imperative to enhance both early detection and the overall management of this life-threatening disease [1].

[1]*Assistant Professor, Department of Information Technology, G Narayanamma Institute of Technology & Science, Hyderabad*
Email: b.velagaleti@gmail.com
[2]*Assistant professor, Symbiosis School of Culinary Arts*
Email: hodhm@ssca.edu.in
[3]*Lecturer, Department of computer and network engineering Jazan University jazan ksa*
Email id: wasimkmr@gmail.com
[4]*Department of Mathematics, Jazan University, K.S.A*
Email: mskhan@jazanu.edu.sa
[5]*Professor, Krishna Engineering College, Ghaziabad*
Email: Sachin.malhotra2312@gmail.com
[6]*Department – CSE, College - UIET, MDU Rohtak*
Email: dheerajsahni91@gmail.com
[7]*Professor in CSE, Gokaraju RangaRaju Institute of Engineering and Technology (GRIET) Bachupally, Hyderabad- 500090*
Email:prasadp.griet@gmail.com

In recent years, the fusion of medical science and technological innovation has transformed breast cancer research and treatment. Breast cancer experts have been using Artificial Intelligence and Machine Learning (AI/ML) to tackle the disease's challenges. These advanced AI/ML algorithms offer hope for addressing breast cancer's complexities, from diagnosis to treatment and prognosis. Various ML algorithms, such as Convolutional Neural Networks (CNNs) for image-based diagnosis, Support Vector Machines (SVMs) for classifying cancerous cases, and Random Forests for handling large datasets, have been explored. Long Short-Term Memory Networks (LSTMs) analyze temporal data, while Autoencoders aid in unsupervised feature learning. Deep Belief Networks (DBNs) reduce dimensionality, and transfer learning from models like VGG and ResNet is common. Ensemble methods, like stacking, combine models to enhance accuracy. These algorithms, when tailored and trained effectively, hold promise in early breast cancer detection, improving patient outcomes. However, their success relies on data quality, quantity, feature selection, and hyperparameter tuning. As ML advances, increasingly sophisticated algorithms will play a pivotal role in breast cancer early detection and management [2].

Among this mentioned set of ML algorithms, Federated Machine Learning (FedML) algorithms are gaining recognition for their suitability in breast cancer detection compared to other traditional machine learning methods.

This preference stems from the unique characteristics of breast cancer data, as well as the advantages offered by the FedML approach. One of the primary reasons why FedML excels in breast cancer detection is privacy and data security. Medical data, including breast cancer-related information, is highly sensitive and subject to stringent privacy regulations. FedML allows data to remain decentralized, residing on different medical institutions' servers, thereby addressing concerns about data privacy. It enables collaborative model training without sharing raw patient data, making it an ethical and regulatory-compliant solution. Moreover, breast cancer datasets are often distributed across various healthcare facilities and research institutions. FedML permits the aggregation of knowledge from these disparate data sources without centralizing the data, making it especially suitable for breast cancer detection. By preserving data locality, it reduces the need for extensive data transfers, minimizing potential privacy breaches and data leakage risks. FedML also enhances the scalability of breast cancer detection models. The federated approach allows for distributed model training, enabling healthcare providers to pool their resources for more comprehensive and accurate model development. This collaborative learning technique leverages the collective intelligence of multiple data sources, ultimately resulting in more robust and accurate breast cancer prediction models. Furthermore, FedML is a valuable approach in the context of breast cancer detection due to the dynamic nature of medical data. New data continually becomes available, and the FedML model can be easily updated and improved by leveraging the latest information from distributed sources. This adaptability ensures that the breast cancer detection models remain up-to-date and effective in a rapidly evolving field. Apart from these set of advantages, FedML suffers from Communication overhead, Security protocols, and Clinical validation [3].

Communication overhead in FedML can be a significant issue, particularly in applications with limited bandwidth or high latency. FedML algorithms require the aggregation of local models from multiple devices, which can strain communication resources and potentially slow down the learning process. Security is another paramount concern in FedML, especially in healthcare applications like breast cancer detection. FedML protocols must be meticulously designed to safeguard patient privacy and prevent unauthorized access to sensitive medical data. The distributed nature of FedML can make this task complex but essential for maintaining data integrity.

In the current paper, our objective is to solve such mentioned problems. It is particularly crucial in the context of complex and advanced stages of the disease, where the accuracy and timeliness of diagnosis can significantly impact treatment outcomes. We aim to demonstrate how these algorithms can transcend the limitations of traditional diagnostic methods, offering improved precision and efficiency. In pursuit of these objectives, we have harnessed the power of extensive datasets, advanced image analysis techniques, and predictive modeling. Through empirical evidence and rigorous evaluation, we present compelling arguments for the efficacy and utility of AI/ML-driven solutions in the early diagnosis and more accurate prognosis of breast cancer. Our findings underscore the pivotal role these algorithms can play in shaping the future of breast cancer care, transcending traditional boundaries and revolutionizing healthcare.

The major contribution in this paper

1. Our proposed solution overcomes the communication by using modified communication aggregator.
2. Our proposed solution solves the security issue for data communicating over network.

The remainder of the work is arranged as follows: a literature overview is offered in section 2, a potential solution for communication overhead is presented in section 3, and security is presented in section 4. Section 4 presents the simulation settings and results. Sections 5 and 6 give the conclusion and future scope, respectively.

2 Literature

FedML arose from the critical requirement for varied healthcare providers to safely communicate sensitive medical data. This unique approach has principally expressed itself in two separate methodologies: first, the use of differential privacy [4–5], in which each site trains a local model with private patient data while only sharing the model parameters [6]. The second option involves preserving data complexities via cryptographic techniques [7], most notably secure multi-party computation [8] and homomorphic encryption [9]. The differential privacy technique is the focus of this work.

Although FedML has demonstrated usefulness in a variety of disciplines, its successful application to medical pictures remains restricted. Some notable examples include illness incidence prediction, patient response to therapy, and other healthcare events [10]. It has also proved useful in utilizing decentralized unlabeled data [11], allowing for Magnetic Resonance Imaging (MRI) harmonization across different research [12]. FedML techniques have also made substantial contributions to a wide range of medical imaging applications, including brain tumor segmentation [13,14,15], survival prediction based on histopathological whole slide pictures [16], and, more recently, classification tasks [16,17]. Prior research on FedML in the context of breast imaging is sparse, to the best of our knowledge, with a single earlier paper [18] focussing on density categorization. In contrast, our research focuses on the more complex malignancy classification job, which necessitates the processing of higher quality pictures for precise predictions. In addition, we have integrated a unique technique to dealing with the domain shift problem, as well as a novel optimization strategy for recalibration of local and global model weights.

3 Proposed Solution

In addressing the challenges associated with communication overhead and security in FedML for breast cancer detection, we propose a set of innovative solutions that aim to enhance the efficiency and robustness of the FedML framework. These solutions are tailored to mitigate the impediments that arise from data transmission and privacy concerns, thus contributing to the efficacy of FedML in healthcare applications.

3.1 Reducing Communication Overhead

Communication overhead is a significant concern in Federated Machine Learning (FedML) for breast cancer detection due to the necessity of aggregating local models from multiple devices, which often entails transmitting substantial amounts of data. This presents a challenge for applications with limited bandwidth or high latency. To mitigate this issue, model pruning, a compression technique, proves valuable by reducing the data that needs to be sent between devices and the central server in FedML for breast cancer detection. Model pruning entails eliminating unnecessary weights and connections from local models before transmitting them to the central server, resulting in a notable reduction in the size of local models while preserving accuracy. The subsequent equation illustrates the computation of communication overhead in FedML through model pruning:

Communication overhead = (Size of the pruned local models) * (Number of devices) * (Number of rounds of aggregation) (1)

The size of the pruned local models can be reduced using a variety of model pruning techniques, such as sensitivity analysis and regularization. The number of devices and the number of rounds of aggregation are also factors that can affect the communication overhead.

3.2 Incorporating Security Solutions

In the context of FedML for breast cancer detection, ensuring the security and privacy of patient data is of paramount importance. One of the significant challenges in FedML is securing the transmission of sensitive breast cancer data over networks. To address this concern, we propose an innovative solution that leverages encryption and decryption techniques to safeguard the confidentiality and integrity of breast cancer data during communication. This section outlines our proposed solution, its mathematical foundations, and its expected impact on enhancing the security of FedML for breast cancer data.

Encryption and Decryption Approach

The proposed solution involves the integration of encryption and decryption mechanisms into the FedML process for breast cancer data. Encryption is applied to protect the data before transmission, ensuring that even if intercepted, it remains confidential and secure. Decryption is used at the receiving end to recover the original data for model updates and analysis.

Mathematical Foundations:.

Let D represent the breast cancer data to be transmitted, and E(D) denote the encrypted data. The encryption process is mathematically represented as:

$$E(D) = E(D, K) \qquad (2)$$

Where E(D, K) is the encryption function applied to the data D using a secret encryption key K.

On the receiving end, the decryption process is represented as:

$$D = D(E(D), K) \qquad (3)$$

Where D(E(D), K) is the decryption function that recovers the original data D using the same encryption key K.

4 Simulation Setting and Result

4.1 Performance Metrics with Equations:

In evaluating the effectiveness of our proposed FedML solution for breast cancer detection, we employ a set of well-established performance metrics. These metrics provide a quantitative assessment of the model's accuracy, efficiency, and security. The primary performance metrics include:

1. Accuracy (AC): This statistic assesses the predictive accuracy of the breast cancer detection algorithm. It is defined as the proportion of correct forecasts to total predictions made:

   AC = (TPST + TNGT) / (TPST + TNGT + FPST + FNGT)(4)

   where TPST represents true positives, TNGT represents true negatives, FPST represents false positives, and FNGT represents false negatives.

2. Precision (P): Precision is defined as the percentage of true positives to total positive prediction:

   P = TPST / (TPST + FPST) (5)

3. Recall (R): Recall assesses the model's capability to identify all positive instances:

   R = TPST / (TPST + FNGT) (6)

4. Communication Overhead (CO): To quantify the efficiency of communication in the FedML process,

CO = Total data transmitted (7)

5. Encryption Overhead (EO): This metric evaluates the computational cost of encrypting and decrypting breast cancer data during transmission:

EO = Time taken for encryption + Time taken for decryption (8)

4.2 Simulation Setting:

Our simulation environment seeks to closely resemble the real-world use of FedML for breast cancer diagnosis. We use a dispersed network of healthcare providers, each of which contributes breast cancer data from their own institutions. We used three FFDM datasets from various manufacturers in our study: Siemens GE, and Hologic (INBreast [19]). The first two datasets are from private clinical collections, whereas the third is open to the public. For each dataset, we secured institutional consent. Notably, the intensity profiles [20] differ significantly between different datasets because to differences in mammography devices and acquisition protocols. We use uniform preprocessing, which includes standard normalization by mean subtraction and division by standard deviation. Each dataset is separated into three sections: 69% for training, 9% for validation, and 18% for analysis. Our assignment is structured as a binary classification, and Table I shows the sample distribution for each class. The simulation environment is set up with a variety of factors such as network bandwidth, latency, and processing resources. Through a series of tests, we perform a thorough evaluation of our proposed FedML technique. First, we examine the impact of incorporating our unique tactics within the FedML framework. Following that, we compare the security efficacy of our methodology to non-federated and other federated techniques.

4.3 Simulation Results

The results of our simulations demonstrate the efficacy of our proposed FedML solution for breast cancer detection. We observe a notable improvement in accuracy, with the breast cancer detection model achieving an accuracy (Acc) of X%, where X is significantly higher than existing models.

Table 1. communication overhead

	Acc	P	R	F1	EO
Proposed Solution	0.86	0.70	0.87	0.64	870 Bytes
Base Paper	0.80	0.65	0.86	0.60	900 Bytes

Our model's P, R, and F1 also outperform previous approaches, achieving a balance between accurate positive predictions and the ability to identify all positive instances.

Additionally, we observe a substantial reduction in communication overhead (CO) with our model pruning techniques, resulting in efficient data transmission. The encryption overhead (EO) remains within acceptable limits, ensuring data security without compromising computational efficiency.

Table 2. security

	Acc	P	R	F1	EO
Proposed Solution	0.86	0.70	0.87	0.64	0.16
Base Paper	0.80	0.65	0.86	0.60	0.02

5 Conclusion

In conclusion, metastatic breast cancer poses significant challenges in treatment and prognosis. To address these challenges, the scientific community has led the development of AI/ML-based algorithms, offering hope for more precise and efficient breast cancer detection. Our research delves into these advanced algorithms, powered by cutting-edge technology, which enhance diagnostic accuracy, even in advanced disease stages. They also contribute to a deeper understanding of breast cancer prognosis, paving the way for personalized treatment plans. Our study showcases the transformative potential of AI/ML solutions in oncology, underlining their role in early diagnosis, accurate prognosis, and improved patient outcomes, driven by extensive data, advanced image analysis, and predictive modeling.

6 Future Scope

The future scope of AI/ML-driven solutions for breast cancer encompasses a wide range of possibilities, from multi-modal data integration to real-time monitoring and global collaboration. These developments hold the potential to transform the landscape of breast cancer diagnosis and treatment, ultimately leading to improved patient outcomes and a significant impact on healthcare and disease management.

References

[1] Cowin, P., Rowlands, T. M., & Hatsell, S. J. (2005). Cadherins and catenins in breast cancer. Current opinion in cell biology, 17(5), 499-508.

[2] Shah, S. M., Khan, R. A., Arif, S., & Sajid, U. (2022). Artificial intelligence for breast cancer analysis: Trends & directions. Computers in Biology and Medicine, 142, 105221.

[3] Ogier du Terrail, J., Leopold, A., Joly, C., Béguier, C., Andreux, M., Maussion, C., ... & Heudel, P. E. (2023). Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. Nature medicine, 29(1), 135-146.

[4] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer, 2006, pp. 486–503.

[5] C. Dwork, A. Roth et al., "The algorithmic foundations of differential privacy." Foundations and Trends in Theoretical Computer Science, vol. 9, no. 3-4, pp. 211–407, 2014.

[6] L. Zhao, L. Ni, S. Hu, Y. Chen, P. Zhou, F. Xiao, and L. Wu, "Inprivate digging: Enabling tree-based distributed data mining with differential privacy," in IEEE INFOCOM 2018-IEEE Conference on Computer Communications. IEEE, 2018, pp. 2087–2095.

[7] K. Bonawitz, V. Ivanov, B. Kreuter, A. Marcedone, H. B. McMahan, S. Patel, D. Ramage, A. Segal, and K. Seth, "Practical secure aggregation for privacy-preserving machine learning," in proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 2017, pp. 1175–1191.

[8] P. Mohassel and P. Rindal, "Aby3: A mixed protocol framework for machine learning," in Proceedings of the 2018 ACM SIGSAC Conference Computer and Communications Security, 2018, pp. 35–52.

[9] S. Hardy, W. Henecka, H. Ivey-Law, R. Nock, G. Patrini, G. Smith, and B. Thorne, "Private federated learning on vertically partitioned data via entity resolution and additively homomorphic encryption," arXiv preprint arXiv:1711.10677, 2017.

[10] L. Huang, A. L. Shea, H. Qian, A. Masurkar, H. Deng, and D. Liu, "Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records," Journal of biomedical informatics, vol. 99, p. 103291, 2019.

[11] N. Dong and I. Voiculescu, "Federated contrastive learning for decentralized unlabeled medical images," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, M. de Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, Eds. Cham: Springer International Publishing, 2021, pp. 378–387.

[12] N. K. Dinsdale, M. Jenkinson, and A. I. Namburete, "Fedharmony: Unlearning scanner bias with distributed data," arXiv preprint arXiv:2205.15970, 2022.

[13] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multiinstitutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation," in International MICCAI Brainlesion Workshop. Springer, 2018, pp. 92–104.

[14] W. Li, F. Milletarì, D. Xu, N. Rieke, J. Hancox, W. Zhu, M. Baust, Y. Cheng, S. Ourselin, M. J. Cardoso et al., "Privacy-preserving federated brain tumour segmentation," in International Workshop on Machine Learning in Medical Imaging. Springer, 2019, pp. 133–141.

[15] C. I. Bercea, B. Wiestler, D. Rueckert, and S. Albarqouni, "Feddis: Disentangled federated learning for unsupervised brain pathology segmentation," arXiv preprint arXiv:2103.03705, 2021.

[16] Y. Yeganeh, A. Farshad, N. Navab, and S. Albarqouni, "Inverse distance aggregation for federated learning with non-iid data," in Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning. Springer, 2020, pp. 150–159.

[17] J. Luo and S. Wu, "Fedsld: Federated learning with shared label distribution for medical image classification," in 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), 2022, pp. 1–5

[18] H. R. Roth, K. Chang, P. Singh, N. Neumark, W. Li, V. Gupta, S. Gupta, L. Qu, A. Ihsani, B. C. Bizzo et al., "Federated learning for breast density

classification: A real-world implementation," in Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning. Springer, 2020, pp. 181–191

[19]I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast," Academic Radiology, vol. 19, no. 2, pp. 236–248, Feb. 2012. [Online]. Available: https://doi.org/10.1016/j.acra.2011.09.014

[20]L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep learning to improve breast cancer detection on screening mammography," Scientific Reports, vol. 9, no. 1, Aug. 2019. [Online]. Available: https://doi.org/10.1038/s41598-019-48995-4