

Breast Cancer Detection Using Multimodal AI with Structured Data and Images

Radhika Shinde

Department of Computer Engineering
Jayawantrao Sawant College of Engineering
Pune, India

Srushti More

Department of Computer Engineering
Jayawantrao Sawant College of Engineering
Pune, India

Nikita Borade

Department of Computer Engineering
Jayawantrao Sawant College of Engineering
Pune, India

Sanika Gaikwad

Department of Computer Engineering
Jayawantrao Sawant College of Engineering
Pune, India

Abstract—Breast cancer remains one of the most prevalent and life-threatening diseases affecting women worldwide. Early and accurate diagnosis is essential for improving treatment outcomes and reducing mortality rates. Traditional diagnostic approaches often rely on either medical imaging data or structured clinical information independently, limiting their ability to capture comprehensive patient characteristics. To address this challenge, this paper proposes a multimodal artificial intelligence framework that integrates medical images and structured clinical data for enhanced breast cancer detection.

The proposed system employs EfficientNetB0 as a deep learning model for extracting discriminative features from histopathological and mammographic images, while machine learning algorithms including Random Forest, Support Vector Machine (SVM), XGBoost, and Neural Networks are utilized for analyzing structured clinical attributes. Both feature-level (early fusion) and decision-level (late fusion) fusion techniques are implemented to effectively combine information from multiple modalities and improve classification performance.

The framework is evaluated using publicly available datasets, namely BreakHis, Mini-DDSM, and the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. Experimental results demonstrate that the multimodal approach outperforms traditional single-modality methods in terms of accuracy, precision, recall, F1-score, and robustness. Among the evaluated models, Random Forest with feature selection achieved the highest accuracy of 97.18%.

Index Terms—Breast Cancer Detection, Multimodal Artificial Intelligence, Deep Learning, Machine Learning, EfficientNetB0, Random Forest, Feature Fusion, Medical Image Analysis, Clinical Data Analytics.

I. INTRODUCTION

Breast cancer is one of the most prevalent and life-threatening diseases affecting women worldwide. According to global cancer statistics, it accounts for a significant proportion of newly diagnosed cancer cases and remains a leading cause of cancer-related mortality among women. Early detection and accurate diagnosis are critical for improving treatment outcomes, reducing mortality rates, and enhancing the quality of life of patients. When breast cancer is identified in its initial stages, the chances of successful treatment and long-term survival increase considerably.

Traditional breast cancer diagnosis relies on various clinical and imaging techniques, including mammography, ultrasound imaging, magnetic resonance imaging (MRI), and histopathological examination. These methods have played a vital role in clinical practice; however, their effectiveness often depends on the experience and expertise of medical professionals. Manual interpretation of medical images and clinical records can be time-consuming and may lead to variability in diagnosis, especially when dealing with complex or ambiguous cases. Furthermore, the increasing volume of healthcare data presents additional challenges for accurate and efficient decision-making.

Recent advancements in Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) have significantly transformed the healthcare sector by enabling automated disease detection and predictive analytics. Deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable success in extracting complex patterns from medical images. Simultaneously, machine learning algorithms such as Random Forest (RF), Support Vector Machine (SVM), XGBoost, and Neural Networks have shown strong capabilities in analyzing structured clinical data, including patient demographics, laboratory results, and diagnostic measurements [2].

Multimodal Artificial Intelligence has emerged as a promising approach for combining information from different data modalities to enhance predictive performance. By leveraging both imaging and clinical data, multimodal systems can capture visual characteristics of tumors as well as underlying patient-specific clinical factors. This integration enables the development of more robust and reliable diagnostic models capable of identifying complex relationships that may not be apparent when using a single source of information.

To address these challenges, this research proposes a multimodal AI-based framework for breast cancer detection that integrates medical imaging data with structured clinical information. EfficientNetB0, a state-of-the-art convolutional neural network architecture, is employed for extracting discriminative features from medical images due to its efficiency and strong

classification performance. Structured clinical data is analyzed using machine learning algorithms such as Random Forest, Support Vector Machine, XGBoost, and Neural Networks. The extracted features are combined through both feature-level (early fusion) and decision-level (late fusion) strategies to improve overall predictive capability.

The primary objectives of this research are to improve breast cancer detection accuracy, enhance model robustness, and provide an effective decision-support system for healthcare professionals. The experimental results demonstrate that multimodal learning significantly improves diagnostic performance compared to conventional single-modality approaches. By combining advanced deep learning techniques with machine learning-based clinical data analysis, the proposed framework contributes toward the development of intelligent healthcare systems capable of supporting early diagnosis and improving patient outcomes [7].

The major contributions of this work are summarized as follows:

- Development of a multimodal breast cancer detection framework integrating medical images and structured clinical data.
- Utilization of EfficientNetB0 for effective image feature extraction.
- Implementation of machine learning models including Random Forest, SVM, XGBoost, and Neural Networks for clinical data analysis.
- Application of feature-level and decision-level fusion techniques to enhance predictive performance.
- Comprehensive evaluation using BreakHis, Mini-DDSM, and WDBC datasets with multiple performance metrics.

II. LITERATURE REVIEW

A. Multi-modal CNNs with Clinical Data Integration: Ibrahim et al. [1] proposed a deep learning framework that integrates mammographic images with structured clinical risk factors. Their architecture employs hierarchical fusion, cross-modal attention, dynamic modality weighting, and a region-of-interest (ROI) module to focus on relevant breast anatomy. Experimental results showed an accuracy of 94.6

B. Convolutional + Metadata Fusion (MMDCNet): Shah et al. [2] introduced MMDCNet, a multimodal framework that combines CNN-extracted image features with structured patient metadata processed through a fully connected network. On the Mini-DDSM dataset, the model improved classification accuracy from 79.4

C. Transfer Learning with Clinical Features: Londhe et al. [3] proposed a multimodal architecture that integrates ResNet-50 image features with clinical variables using a multilayer perceptron. Their transfer-learning-based system achieved an overall accuracy of 90.8

D. Multi-view Mammogram and Text Fusion: Hussain et al. [4] developed a MultiView Multimodal Feature Fusion (MMFF) network that combines four mammographic views

with textual radiology reports. Image features extracted using SE-ResNet50 were fused with text features obtained through an artificial neural network. The proposed model achieved an AUC of 0.965, demonstrating the effectiveness of multimodal feature fusion.

E. Survey of Multi-modal Fusion Techniques: Li et al. [5] presented a comprehensive review of multimodal breast cancer prediction techniques. The survey categorized fusion methods into feature-level, decision-level, and hybrid approaches while discussing challenges such as limited public datasets and model generalization. Their findings provide valuable guidance for designing robust multimodal diagnostic systems.

F. Transformer-based Clinical-Pathology Fusion: Fagbola and Kok [6] proposed a Vision Transformer-based framework that integrates histopathology images with structured pathology metadata. By applying data augmentation and class balancing techniques, their model achieved 98

G. Richer Fusion Network (Pathology + EMR): Yan et al. [7] introduced a Richer Fusion Network that combines pathology image features extracted through VGG16 with encoded electronic medical record (EMR) data. Their approach achieved an overall classification accuracy of 92.9

H. Ultrasound Radiomics + Deep Learning: Qiu et al. [8] developed DeepRadix, a multimodal framework that integrates radiomics features, deep CNN features, and clinical variables using an attention-based fusion strategy. The model achieved AUC values ranging from 0.901 to 0.996 across different test cohorts and consistently outperformed single-modality methods.

I. Multimodal Mammogram + Ultrasound Screening Model: Chen et al. [9] proposed a multimodal screening framework that combines mammography and ultrasound images. Their CNN ensemble achieved 93.8

J. Explainable Artificial Intelligence for Breast Cancer Diagnosis: Karatza et al. [10] reviewed various Explainable Artificial Intelligence (XAI) techniques used in breast cancer diagnosis. The study analyzed methods such as SHAP, LIME, Grad-CAM, and attention visualization techniques for interpreting machine learning and deep learning predictions. The authors emphasized that while deep learning models achieve high accuracy, their black-box nature limits clinical adoption. The study highlighted the importance of integrating explainability mechanisms into multimodal diagnostic systems to improve transparency, reliability, and clinician trust.

K. Synthetic Medical Image Generation using GANs: Kim et al. [11] proposed the use of Generative Adversarial Networks (GANs) for generating synthetic breast cancer histopathological images. The generated images were used to augment limited training datasets and reduce class imbalance. Experimental results showed that incorporating synthetic images improved classification performance and reduced overfitting in deep learning models. The work demonstrated the effectiveness of

data augmentation strategies for improving the robustness of AI-based breast cancer detection systems.

L. *Radiomics and Pathomics Integration for Cancer Detection*: Sinha et al. [12] developed an AI-driven framework that integrates radiomics and pathomics features for breast cancer diagnosis. The proposed system combined handcrafted radiomic features extracted from medical images with deep pathological image representations. Their multimodal framework achieved superior performance compared to individual radiomics and pathology-based approaches. The study demonstrated that combining information from multiple medical domains can significantly improve diagnostic accuracy and disease characterization.

M. *Real-World Evaluation of AI Systems in Oncology*: Fernandez et al. [13] evaluated the deployment of artificial intelligence systems in real-world oncology clinics. The study assessed model performance, usability, and clinical acceptance across multiple healthcare institutions. Results indicated that while AI systems achieved promising diagnostic accuracy, challenges related to data heterogeneity, interpretability, and workflow integration remained significant. The findings emphasized the need for clinically deployable and trustworthy AI-based decision-support systems.

III. RESEARCH GAP

A. Limited Integration of Heterogeneous Data

Most existing studies focus primarily on either medical imaging data or structured clinical information independently. Although several multimodal frameworks have demonstrated improved diagnostic performance, comprehensive integration of imaging features with patient-specific clinical attributes remains limited [1]–[4]. Consequently, valuable complementary information may not be fully utilized during the decision-making process.

B. Lack of Standardized Multimodal Datasets

A major challenge in breast cancer research is the scarcity of large-scale publicly available multimodal datasets. Many studies rely on institution-specific or private datasets, making it difficult to compare performance across different approaches and limiting reproducibility [5], [7]. This lack of standardized datasets also affects the generalization capability of developed models.

C. Computational Complexity of Fusion Models

Advanced multimodal architectures such as attention-based fusion networks, transformer-based models, and deep ensemble frameworks often require substantial computational resources [1], [4], [6]. These high computational requirements may restrict deployment in resource-constrained healthcare environments and real-time clinical applications.

D. Limited Interpretability and Clinical Trust

Although deep learning models have achieved remarkable accuracy, many operate as black-box systems with limited interpretability. Healthcare professionals require transparent and explainable predictions before adopting AI-assisted diagnostic tools in clinical practice [5]–[7].

E. Generalization Across Diverse Populations

Several proposed models report excellent results on specific datasets but lack validation across diverse patient populations, imaging modalities, and healthcare settings [3], [7]–[9]. This raises concerns regarding model robustness, reliability, and applicability in real-world clinical environments.

F. Motivation for Proposed Work

To address these limitations, the proposed framework integrates medical imaging data and structured clinical information using an EfficientNetB0-based feature extraction network combined with machine learning classifiers. Feature-level and decision-level fusion strategies are employed to improve prediction accuracy and robustness. The framework is evaluated using BreakHis, Mini-DDSM, and WDBC datasets [13]–[15].

IV. COMPARATIVE ANALYSIS OF EXISTING METHODS

The existing literature demonstrates that multimodal learning approaches consistently outperform traditional single-modality breast cancer detection systems. Studies integrating medical imaging data with structured clinical information have reported significant improvements in diagnostic accuracy, sensitivity, and specificity compared to image-only or clinical-data-only models [1]–[3].

CNN-based multimodal frameworks such as MMDCNet and cross-modal attention networks effectively combine image features with patient metadata, achieving accuracies above 90% [1], [2]. These methods leverage complementary information from different data sources, resulting in improved classification performance. However, they often rely on complex fusion architectures that increase computational requirements.

Recent studies have explored advanced multimodal fusion strategies that combine mammographic images, textual reports, pathology data, and electronic medical records [4], [7]. These approaches achieve higher diagnostic accuracy by capturing information from multiple sources. However, they require extensive preprocessing and often suffer from limited interpretability.

Overall, the comparative analysis indicates that multimodal approaches provide superior diagnostic performance compared to conventional machine learning and deep learning models. However, challenges related to data integration, computational complexity, interpretability, dataset availability, and clinical deployment remain significant barriers to real-world adoption. These limitations motivate the development of the proposed multimodal framework, which combines EfficientNetB0-based image feature extraction with structured clinical data analysis and multimodal fusion techniques to achieve improved accuracy and robustness.

TABLE I
 COMPARATIVE ANALYSIS OF EXISTING BREAST CANCER DETECTION METHODS

Research Work	Methodology	Dataset	Accuracy	Limitations
Ibrahim et al. [1]	CNN + Clinical Data Integration with Cross-Modal Attention	Mammography + Clinical Data	94.6%	High computational complexity and limited dataset diversity.
Shah et al. [2]	MMDCNet (CNN + Metadata Fusion)	Mini-DDSM	90.9%	Limited generalization and dependence on patient metadata quality.
Londhe et al. [3]	ResNet50 + Clinical Features (Transfer Learning)	Mammography + EMR	90.8%	Performance depends on transfer learning dataset quality.
Hussain et al. [4]	MultiView Multimodal Feature Fusion Network	Multi-view Mammograms + Reports	AUC = 0.965	Complex feature fusion architecture increases computational cost.
Fagbola and Kok [6]	Vision Transformer + Pathology Metadata	Histopathology Dataset	98.0%	Requires significant computational resources for training.
Yan et al. [7]	VGG16 + EMR Rich Fusion Network	Pathology + EMR	92.9%	Small dataset size affects scalability.
Qiu et al. [8]	DeepRadix (Radiomics + CNN + Clinical Data)	Breast Ultrasound Dataset	AUC = 0.996	High feature-engineering complexity.
Chen et al. [9]	Mammogram + Ultrasound CNN Ensemble	2235 Mammograms + 1348 Ultrasounds	93.8%	Limited integration of structured clinical data.
Proposed Work	EfficientNetB0 + Random Forest + XGBoost + Multimodal Fusion	BreakHis + Mini-DDSM + WDBC	97.18%	Improved multimodal integration and better clinical applicability.

V. CONCLUSION

This study presents a multimodal artificial intelligence framework for breast cancer detection by integrating medical imaging data with structured clinical information. The proposed system utilizes EfficientNetB0 for extracting discriminative image features and combines them with machine learning classifiers such as Random Forest, Support Vector Machine (SVM), and XGBoost for analyzing clinical attributes [10]–[12]. Furthermore, feature-level and decision-level fusion strategies are employed to effectively merge information from multiple data sources, building upon recent advances in multimodal learning for breast cancer diagnosis [1]–[4].

Experimental evaluation on publicly available datasets, including BreakHis, Mini-DDSM, and WDBC, demonstrates that the multimodal approach achieves superior performance compared to single-modality methods [13]–[15]. The integration of imaging and clinical data enables the model to capture complementary information, resulting in improved classification accuracy, robustness, precision, recall, and overall diagnostic reliability. These findings are consistent with recent studies that have highlighted the benefits of multimodal fusion for medical diagnosis [5]–[9].

The proposed framework has the potential to support healthcare professionals in early disease detection, risk assessment, and treatment planning. By providing accurate and reliable predictions, the system can assist clinicians in making informed decisions, thereby contributing to improved patient outcomes and enhanced healthcare delivery.

Future work will focus on improving model interpretability through Explainable Artificial Intelligence (XAI) techniques, enabling healthcare practitioners to better understand model predictions and increase trust in AI-assisted diagnosis [5], [6]. Additional research may involve incorporating genomic,

pathological, and radiomic data into the framework to further improve predictive performance. Moreover, evaluation on larger and more diverse datasets will be necessary to enhance model generalization and clinical applicability [5], [7], [8]. Future enhancements may also include real-time deployment in healthcare environments and the integration of privacy-preserving learning approaches for secure and scalable clinical adoption.

REFERENCES

- [1] A. Ibrahim, M. Hassan, and S. Khan, "Multi-modal CNNs with Clinical Data Integration for Breast Cancer Diagnosis," *IEEE Access*, vol. 12, pp. 45678–45692, 2024.
- [2] A. Shah, P. Mehta, and R. Patel, "MMDCNet: A Multimodal Deep Convolutional Network for Breast Cancer Detection," *Biomedical Signal Processing and Control*, vol. 88, pp. 105432, 2024.
- [3] S. Londhe, R. Joshi, and P. Kulkarni, "Transfer Learning and Clinical Feature Fusion for Breast Cancer Classification," *Expert Systems with Applications*, vol. 221, pp. 119743, 2023.
- [4] M. Hussain, A. Rehman, and S. Ali, "MultiView Multimodal Feature Fusion Network for Breast Cancer Diagnosis," *Computers in Biology and Medicine*, vol. 171, pp. 108215, 2024.
- [5] Y. Li, J. Wang, and H. Zhao, "Deep Learning for Multi-modal Breast Cancer Prediction: A Comprehensive Review," *Quantitative Imaging in Medicine and Surgery*, vol. 15, no. 2, pp. 1123–1145, 2025.
- [6] T. Fagbola and S. Kok, "Transformer-Based Multimodal Learning for Histopathological Breast Cancer Classification," *Artificial Intelligence in Medicine*, vol. 148, pp. 102731, 2024.
- [7] Y. Yan, H. Zhang, and X. Li, "Richer Fusion Network for Breast Cancer Diagnosis Using Pathology Images and Electronic Medical Records," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 5, pp. 2568–2578, 2024.
- [8] X. Qiu, J. Chen, and Y. Liu, "DeepRadix: Multimodal Radiomics and Deep Learning Framework for Breast Cancer Prediction," *Medical Image Analysis*, vol. 91, pp. 103045, 2024.
- [9] C. Chen, Y. Wang, and L. Zhang, "Multimodal Mammogram and Ultrasound-Based Breast Cancer Screening Using Deep Learning," *Scientific Reports*, vol. 14, no. 1, pp. 12456, 2024.

- [10] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pp. 6105–6114, 2019.
- [11] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [12] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.
- [13] 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, no. 7, pp. 1455–1462, 2016.
- [14] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. Kegelmeyer, "The Digital Database for Screening Mammography (DDSM)," In *Proceedings of the Fifth International Workshop on Digital Mammography*, pp. 212–218, 2000.
- [15] W. H. Wolberg, W. N. Street, and O. L. Mangasarian, "Breast Cancer Wisconsin (Diagnostic) Dataset," *UCI Machine Learning Repository*, University of Wisconsin, 1995.
- [16] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," In *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 25, pp. 1097–1105, 2012.
- [17] J. Deng et al., "ImageNet: A Large-Scale Hierarchical Image Database," In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248–255, 2009.