Literature Review

Breast cancer is the most frequently diagnosed cancer among women worldwide, remaining a primary cause of mortality, especially in resource-limited areas (Jung et al., 2013). In 2020, approximately 2.3 million new cases of breast cancer were recorded, resulting in an estimated 684,000 deaths globally (Cancer Today, 2020). The development of breast cancer is influenced by a multitude of risk factors, including age, genetic susceptibility, familial cancer history, and lifestyle variables such as diet, obesity, and exposure to environmental pollutants (Michael et al., 2022). Typically originating in the ducts or lobules of the breast, breast cancer often progresses gradually in its initial stages before potentially metastasizing (Bernardi et al., 2012). Although many breast lumps are benign, malignant tumors can develop asymptomatically, underscoring the critical need for routine screenings, particularly given the absence of early symptoms. Early intervention is pivotal for favorable outcomes, as untreated breast cancer tends to progress, complicating treatment and diminishing survival rates (Sung et al., 2021).

The evolution of medical imaging technologies has significantly enhanced breast cancer diagnostics, enabling earlier and more precise interventions. Mammography remains the principal screening modality for breast cancer, with extensive research affirming its capacity to reduce mortality rates by enabling early detection, especially among women aged 40 to 74 (López et al., 2021). Digital mammography has further improved diagnostic accuracy by providing high-resolution imaging, which enhances visualization of overlapping tissues and reduces false negatives (Mall et al., 2017). Digital Breast Tomosynthesis (DBT), an innovative 3D mammography technology, captures layered views of breast tissue, facilitating the differentiation of benign and malignant structures (Dhungel, Carneiro and Bradley, 2015). Recent studies indicate that DBT offers heightened sensitivity and specificity compared to traditional mammography, especially for patients with dense breast tissue, thereby enabling the detection of smaller, early-stage malignancies.

In addition to mammography, supplementary imaging techniques such as ultrasound and Magnetic Resonance Imaging (MRI) have become indispensable in breast cancer diagnosis. Ultrasound, which uses high-frequency sound waves to create breast images without ionizing radiation, is particularly useful for dense breast tissue and for screening younger women at elevated risk (Jiang et al., 2024). However, ultrasound's dependency on the operator can affect the consistency of results, and it generally lacks the specificity required for independent screening. MRI, with its high sensitivity for detecting minute tissue abnormalities, has proven efficacious in identifying invasive breast cancers in women with genetic predispositions, providing detailed visualizations of soft tissues (Alotaibi et al., 2023). However, MRI's high cost and limited accessibility constrain its widespread application in general population screening, making it more appropriate for high-risk individuals.

Emergent imaging modalities, such as Diffuse Optical Tomography (DOT), present promising non-invasive, radiation-free alternatives for breast cancer imaging. DOT employs near-infrared light to differentiate tissue properties, including oxygenation and hemoglobin concentration, which may indicate malignant tissue presence (Kuttan and Elayidom, 2023). Despite its potential, DOT faces substantial computational challenges in image reconstruction and is currently less established than mammography and MRI. Ongoing research seeks to optimize DOT, potentially using it in conjunction with established imaging modalities to enhance sensitivity and specificity in breast cancer diagnostics.

Techniques	Sensitivity	Tumour size correspondin g to sensitivity	Advantages	Disadvantages
Mammography	85	≤2 cm	Improved image resolution, widely available	Limited sensitivity in dense breast tissue, exposure to radiation
Ultrasound	82	2cm	No ionizing radiation, suitable for dense breasts and implant imaging	Operator-dependent, limited specificity
MRI	95	<2cm	mages small details of soft tissues	Expensive
Diffused Optical Tomography	92	1cm	mages small details of soft tissues	Ill-posed problem during reconstruction

The availability of extensive, well-annotated mammography datasets has been instrumental in advancing breast cancer diagnostics, particularly in fostering artificial intelligence (AI) and machine learning-based methodologies. The Mammographic Image Analysis Society (MIAS) dataset, established in 1994, contains labeled mammograms of both benign and malignant lesions, serving as an early cornerstone for automated breast cancer detection research (Suckling et al., 1994). The Digital Database for Screening Mammography (DDSM), developed in the early 2000s, comprises thousands of mammogram images annotated by experts, significantly aiding the development of computer-aided detection (CAD) systems (Heath et al., 2001). CAD systems trained on these datasets have demonstrated notable improvements in detecting microcalcifications and masses, expediting diagnosis and enhancing accuracy.

More contemporary datasets, such as INbreast (2012), provide high-quality full-field digital mammography (FFDM) images, reflecting current clinical standards and proving instrumental for training convolutional neural networks (CNNs) and other machine learning models in breast cancer detection (Moreira et al., 2012). The INbreast dataset offers detailed annotations on breast density, mass shape, and calcifications, facilitating the training of models to recognize nuanced cancer indicators. CBIS-DDSM (Curated Breast Imaging Subset of DDSM), developed through the National Cancer Institute's Cancer Imaging Archive, includes digital images and has become a valuable resource for training advanced deep learning models (Lee et al., 2017). The availability of these comprehensive datasets has enabled further innovations in machine learning, thereby improving detection rates and reducing false positives.

Computer-aided detection (CAD) systems have become essential in breast cancer diagnosis by alleviating radiologists' workload and enhancing diagnostic accuracy. Typically, CAD systems involve stages of image preprocessing, feature extraction, feature selection, and classification (Godavarty et al., 2015). Image preprocessing techniques improve image clarity, allowing for better visualization of potential tumor regions by minimizing noise (Singh et al., 2020). The detection of regions of interest (ROIs), often achieved through techniques like fuzzy logic and clustering, is critical in identifying suspicious areas within images for further analysis (Wang et al., 2019). Recent advancements in image preprocessing and feature extraction methods have allowed CAD systems to analyze images with higher precision, resulting in more accurate identification of abnormal tissue structures.

Feature extraction in CAD systems captures key image characteristics, such as texture, shape, and intensity, enabling classifiers to distinguish between benign and malignant regions. Traditional machine learning algorithms, like support vector machines (SVMs) and decision trees, have achieved high classification accuracy in identifying malignant lesions (Dhungel et al., 2017). However, these approaches are constrained by their reliance on manually engineered features, which may not fully capture the complexity of mammographic images. Modern CAD systems harness deep learning to automate feature extraction and classification, enhancing diagnostic accuracy and efficiency (Zhu et al., 2019).

Deep learning, especially through convolutional neural networks (CNNs), has revolutionized breast cancer diagnostics by significantly improving feature extraction and classification capabilities. CNNs utilize large datasets to learn hierarchical representations of medical images, refining their capacity to identify complex patterns indicative of cancer. Advanced CNN architectures, such as ResNet, DenseNet, and Inception, have further enhanced diagnostic accuracy, as these models use residual connections and multi-scale feature extraction to identify fine-grained details within images (Litjens et al., 2017). Studies by Esteva et al. (2019) have demonstrated that CNN-based models can achieve accuracy levels comparable to radiologists in identifying malignancies, a promising development for breast cancer diagnostics.

Transfer learning, a method in which pre-trained models are adapted for new tasks, has proven especially beneficial in medical imaging, where labeled data is scarce. Transfer learning enables researchers to use models trained on extensive general datasets, adapting them for specific tasks like breast cancer detection (Fletcher and Elmore, 2013). By leveraging transfer learning, deep learning models achieve enhanced accuracy with less annotated data, making this approach particularly valuable in medical applications.

Optimization of hyperparameters, such as learning rate and batch size, is also crucial for fine-tuning deep learning models to improve diagnostic accuracy in breast cancer detection. Studies have indicated that careful hyperparameter tuning can significantly increase model performance, as demonstrated by (Litjens et al., 2017) Recent research has also focused on feature reduction techniques, which streamline model efficiency by discarding redundant features, thereby improving classification precision (Wellings, Vassiliades and Abdalla, 2016). Incorporating feature selection algorithms and dialectical frameworks into deep learning further enhances robustness and accuracy.

To evaluate AI model performance in breast cancer detection, various metrics are used, including the Dice Similarity Index, Jaccard Coefficient, and F1 Score. These metrics provide critical insights into model accuracy, reliability, and predictive power. Studies employing advanced models like Mask R-CNN and MS-ResCU-Net have achieved high sensitivity and specificity in segmentation tasks, indicating the potential for these systems in clinical applications (Global Burden of Disease Cancer Collaboration et al., 2015) However, real-world challenges persist, including model interpretability, patient-specific variability, and the need for further validation across diverse patient populations

Year	Name	Evaluation Dataset	Noise Removal Method	Performance Metrics (Results)
2015	CRF	INbreast, DDSM-BCRP	N/A	89% Dice Index
2018	Adversarial Deep Structured Net	INbreast, DDSM-BCRP	N/A	97.0% segmentation rate
2018	Deep Learning using YOLO	INbreast	NA	Detection rate of 98.96%, MCC of 97.62%, F1 score of 99.24%
2019	MS-ResCU-Net and ResCU-Net	INbreast	NA	Accuracy: 94.16%, Sensitivity: 93.11%, Specificity: 95.02%, Dice index: 91.78%, Jaccard: 85.13%, MCC: 87.22%
2020	U-Net	CBIS-DDSM, INbreast	Adaptive median filter	Mean Dice coefficient index of 95.10%, mean IOU of 90.90%
2020	Mammographic CAD based on pseudocolor and Mask RCNN	INbreast	Morphological filters	Dice Similarity Index of 0.88, True Positive Rate (TPR) of 0.90
2023	DenseNet with Mask-CNN for Feature Selection	INbreast	NA	>90% accuracy in tumor classification

The integration of advanced imaging technologies and AI has catalyzed significant improvements in breast cancer detection and diagnostics. While mammography remains widely used, its limitations in dense breast tissue underscore the necessity for multimodal imaging approaches. Emerging techniques, such as DBT, MRI, and DOT, show promise for more comprehensive and accurate diagnostics, especially when combined with AI methodologies.

AI, and particularly deep learning, has shown substantial promise in enhancing diagnostic accuracy and minimizing human error in breast cancer detection. Nonetheless, challenges such as data quality, class imbalance, and interpretability remain. Further research may focus on developing hybrid models that integrate AI with conventional imaging and personalized medical data, paving the way for more precise, individualized screening protocols. Future advancements in dataset quality, model interpretability, and interdisciplinary collaboration will be essential to unlocking AI's full potential in improving breast cancer outcomes and survival rates.

References

1. Jung, R.E., Fenner, N., Hämmerle, C.H. and Zitzmann, N.U. (2013) 'Long-term outcome of implants placed with guided bone regeneration (GBR) using resorbable and non-resorbable membranes after 12-14 years', *Clinical Oral Implants Research*, 24(10), pp. 1065–1073. doi: 10.1111/j.1600-0501.2012.02522.x.
2. American Cancer Society, n.d. *Breast Cancer*. Available at: https://www.cancer.org.
3. Michael, E., Ma, H., Li, H. and Qi, S. (2022) 'An optimized framework for breast cancer classification using machine learning', *Biomedical Research International*, 2022, Article ID 8482022. doi: 10.1155/2022/8482022.

4. Bernardi, D., Ciatto, S., Pellegrini, M., Anesi, V., Burlon, S., Cauli, E., Depaoli, M., Larentis, L., Malesani, V., Targa, L., Baldo, P. and Houssami, N. (2012) 'Application of breast tomosynthesis in screening: incremental effect on mammography acquisition and reading time', *British Journal of Radiology*, 85(1020), pp. e1174–e1178. doi: 10.1259/bjr/19385909.
5. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A. and Bray, F. (2021) 'Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries', *CA: A Cancer Journal for Clinicians*, 71(3), pp. 209–249. doi: 10.3322/caac.21660.
6. Godavarty, A., Rodriguez, S., Jung, Y.J. and Gonzalez, S. (2015) 'Optical imaging for breast cancer prescreening', *Breast Cancer (Dove Medical Press)*, 7, pp. 193–209. doi: 10.2147/BCTT.S51702.
7. López, N.C., García-Ordás, M.T., Vitelli-Storelli, F., Fernández-Navarro, P., Palazuelos, C. and Alaiz-Rodríguez, R. (2021) 'Evaluation of feature selection techniques for breast cancer risk prediction', *International Journal of Environmental Research and Public Health*, 18(20), p. 10670. doi: 10.3390/ijerph182010670.

8. Mall, S., Lewis, S., Brennan, P., Noakes, J. and Mello-Thoms, C. (2017) 'The role of digital breast tomosynthesis in the breast assessment clinic: a review', *Journal of Medical Radiation Sciences*, 64(3), pp. 203–211. doi: 10.1002/jmrs.230.

9. Dhungel, N., Carneiro, G. and Bradley, A.P. (2015) 'Automated mass detection in mammograms using cascaded deep learning and random forests', in *Proceedings of the 2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, Adelaide, SA, Australia, 2015, pp. 1–8. doi: 10.1109/DICTA.2015.7371234.

10. Jiang, B., Bao, L., He, S. et al. (2024) 'Deep learning applications in breast cancer histopathological imaging: diagnosis, treatment, and prognosis', *Breast Cancer Research*, 26, p. 137. Available at:

11. Alotaibi, M., Aljouie, A., Alluhaidan, N., Qureshi, W., Almatar, H., Alduhayan, R., Alsomaie, B. and Almazroa, A. (2023) 'Breast cancer classification based on convolutional neural network and image fusion approaches using ultrasound images', *Heliyon*, 9(11), e22406. doi: 10.1016/j.heliyon.2023.e22406.

12. Kuttan, G.O. and Elayidom, M.S. (2023) 'Review on computer aided breast cancer detection and diagnosis using machine learning methods on mammogram image', *Current Medical Imaging*, 19(12), pp. 1361–1371. doi: 10.2174/1573405619666230213093639.
13. Sawyer-Lee, R., Gimenez, F., Hoogi, A., & Rubin, D. (2016). **Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) [Data set]**. The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.7O02S9CY
14. Wellings E, Vassiliades L, Abdalla R. Breast Cancer Screening for High-Risk Patients of Different Ages and Risk - Which Modality Is Most Effective? Cureus. 2016 Dec 28;8(12):e945. doi: 10.7759/cureus.945. PMID: 28133583; PMCID: PMC5268380.

15. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B. and Sánchez, C.I. (2017) ‘A survey on deep learning in medical image analysis’, *Medical Image Analysis*, 42, pp. 60–88. doi: 10.1016/j.media.2017.07.005.
16. Fletcher, S.W. and Elmore, J.G. (2003) ‘Clinical practice. Mammographic screening for breast cancer’, *New England Journal of Medicine*, 348(17), pp. 1672–1680. doi: 10.1056/NEJMcp021804.
17. INbreast (n.d.) *INbreast: A full-field digital mammography database for breast cancer detection*. Available at: https://www.inbreast.org (Accessed: 11 November 2024).
18. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S. and Dean, J. (2019) ‘A guide to deep learning in healthcare’, *Nature Medicine*, 25(1), pp. 24–29. doi: 10.1038/s41591-018-0316-z.