Grenze International Journal of Engineering and Technology, June Issue

Breast Cancer Detection using Histopathology Images

Pawaskar Mayuri Ashok[1] and K. S. Gandle[2]
[1-2]Computer Science and Engineering, Government College of Engineering Karad, India
Email: pawaskarmayuri2095@gmail.com

***Abstract*— One of the leading causes of death among women in the world is breast cancer and to make sure that the treatment process is effective and the survival rates could rise, the state of the disease should be diagnosed in time and rightly. The manual analysis of histopathology images under the scope of viewing them through a manual is excruciatingly time-consuming and prone to human error. In the recent years, the use of AI and DL methods can be used to automatize the process of diagnostics and ensure positive outcomes. The study will be focused on developing a deep learning model that will recognize and classify breast cancer histopathology images. The algorithm of the spatial and textual features extraction of the images is performed by the means of convolutional neural networks (CNNs), thus, it does not involve any hand-crafted feature engineering. The model is trained and tested on publicly available datasets, besides being able to make recommendations with considerable confidence about the diagnosis of benign and malignant specimens. The model provides diagnostic support to the pathologists that is quick and repeatable as well as consistent. As an example of potential of DL, the current paper discusses analysis of images as a supplement to computer-aided diagnosis of breast cancer and the automation of pathological processes.**

***Index Terms*— Breast cancer detection, Histopathology images, Computer-aided diagnosis (CAD) Digital pathology.**

I. INTRODUCTION

Breast cancer continues to be one of major diseases that are common and could lead to the death of women around the globe. The World Health Organization (WHO) reports that millions of new cases are recorded annually and the disease remains the main cause of cancer-related death in women. Therefore, early detection and precise diagnosis are the keys to patient outcome improvements and survival rates growth. In the past, visually looking at histopathology images of tissues under a microscope by experienced pathologists to find cancerous changes made up the breast cancer diagnostic method. Yet, such a manual process takes a long time, is subjective, and has the problem of variability between different observers. However, with the help of AI and DL, CAD systems in detection of cancer have become leading tools that can converse with pathologists and seek their assistance in the detection. DL models, especially CNNs, have attained great achievements in image classification and pattern recognition challenges. Thus, CNNs in histopathology images can independently initiate the most metabolic pathways that certainly differentiate characteristics of the tissues that are malign or benign without the intervention of feature extraction techniques from the human side. This work is to figure out the creation of a ML based model for breast cancer detection using histopathology images. The essential point is the elevation of diagnostic accuracy, the ramping down of the analysis time, and the facilitation of the medical professionals to decide in an unbiased and equilibrated way.

Grenze ID: 01.GIJET.12.2.6

The adopted method has been inculcated and checked through publicly available histopathological datasets, and standard performance evaluation metrics like accuracy, precision, recall, and F1-score have been utilized to measure results. Harmonizing of such smart systems with clinical practice environment can be compared to the impact that the advent of the airplane had on the mode of long-distance travel. It has the potential to make breast cancer diagnosis fast, dependable, and easy to repeat.

Traditional Diagnostic Techniques

Mammography

Mammography is by far most common method used for early breast cancer screening. It involves use of low dose X-rays to take pictures of breast tissue and detect any abnormalities such as lumps or calcifications. However, the issue of false negatives or false positives might arise in a younger woman with dense breast tissue, where mammography has certain limitations.

Ultrasound Imaging

Breast ultrasound imaging is non invasive technique that uses high frequency sound waves to generate accurate images of breast tissues. Commonly performed alongside mammography to differentiate solid masses (possible tumors) from fluid-filled cysts. Nevertheless, the skills of the interpreting radiologist are critical in the ultrasound diagnosis.

Magnetic Resonance Imaging (MRI)

By the use of an MRI, detailed, three-dimensional, high-resolution images of the breast can be obtained. This method is extremely valuable in identifying tiny lesions and for the evaluation of cancer that has spread to other areas. However, due to several reasons such as expensive, time-consuming, and not suitable for routine screening, MRI should not be used in regular screening.

Histopathological Examination

Histopathology is considered the most reliable method for breast cancer confirmation. It based on microscopic investigation of the biopsy samples stained with Hematoxylin and Eosin (H\&E). The pathologists look at cells, their structures, forms, and patterns to figure out if the sample is cancerous or not. Although, the manual inspection is subjective, slow, and has human errors.

TABLE I. COMPARISON OF TECHNIQUES FOR BREAST CANCER DETECTION

Technique	Application	Limitations
Mammography	Used for early screening and detection of breast tumors using low-dose X-rays.	Limited accuracy in dense breast tissue; possible false positives and negatives.
Ultrasound Imaging	Helps differentiate between solid tumors and cysts; often used after mammography.	Operator-dependent; less effective in detecting small or early-stage tumors.
MRI	Provides detailed 3D visualization of breast tissue; used for assessing cancer spread.	High cost and long scanning time; not suitable for routine screening.
Histopathological Examination	Microscopic analysis of biopsy samples to confirm presence of malignant cells.	Manual and time-consuming; prone to observer variability and human error.
Clinical Breast Examination (CBE)	Performed by healthcare professionals to identify lumps or abnormalities.	Subjective and depends on examiner skill; cannot detect deep or small tumors.

A. Key Contributions of the research

- A deep learning–based framework is developed for automated detection and classification of breast cancer from histopathology images, reducing dependence on manual examination.
- Advanced preprocessing techniques, including stain normalization and patch extraction, are applied to enhance image consistency and improve feature representation.
- A customized CNN architecture is designed and optimized to effectively capture high-resolution tissue patterns, achieving improved diagnostic performance.
- Explainable AI techniques are integrated to provide visual insights into model predictions, supporting transparency and aiding pathologists in decision-making.

II. LITERATURE REVIEW

Lekha S. Nair, K. R. Amarnath, and Jyothisha J. Nair (2025) proposed a deep learning-based model that focuses on improving accuracy and efficiency of breast cancer classification using histopathology images. The authors introduced a hybrid SE-Conformer framework that integrates Squeeze-and-Excitation Residual Convolutional Blocks (SE-Res-Conv) with a Conformer block, enabling the model to capture both local and global dependencies for more precise feature extraction and classification. Tested on the BreakHis and BACH datasets,

model achieved high accuracy of 97 on 200× magnification images, outperforming existing CNN and transformer-based methods. The main contribution lies in designing a lightweight yet powerful convolution-attention architecture that enhances feature representation and reduces computational complexity. The authors conclude that their model can significantly aid computer-aided diagnosis systems and plan to extend future work toward real-time clinical validation and integration with explainable AI for improved interpretability in medical applications [1].

Elsheakh, D. N. \& et.al (2024) proposed an innovative approach for early breast cancer detection. The study introduces a microwave-based wearable detection system embedded in Smart Bra that integrates flexible coplanar waveguide (CPW) monopole antennas as sensors operating in the 1.5–8 GHz range. These sensors detect malignant tumors by analyzing scattering parameters (S11, S21, etc.) obtained from multiple sensor configurations around the breast. To enhance tumor identification accuracy, the authors applied machine learning algorithms, particularly the CatBoost classifier, which achieved up to 98\% detection accuracy for tumor sizes as small as 5 mm using a 2×2 antenna sensor array. The main contribution lies in designing a safe, conformal, and non-invasive flexible sensing system capable of real-time breast cancer monitoring while maintaining low specific absorption rate (SAR) values within safety limits. The experimental validation using realistic rubber breast phantoms demonstrated the system's reliability and localization ability. The authors suggest that future work will focus on integrating artificial intelligence and embedded electronics into the Smart Bra for real-time diagnosis, improving miniaturization, and conducting clinical validation studies to move toward widespread biomedical application [2].

Aratake, S.et.al (2023) conducted a study titled "Behavior of Breast Cancer Cells under Oxygen Concentration Gradients in Microfluidic Device" to investigate how varying oxygen levels affect breast cancer cell behavior. They developed a microfluidic device capable of generating controlled oxygen concentration gradients that mimic the hypoxic conditions found in the tumor microenvironment (TME). They analyzed cell proliferation, migration, and death under different oxygen gradients (0–5 , 0–10 , and 0–21). The findings revealed that cancer cells exhibited increased proliferation around 5 oxygen concentration, indicating that moderate hypoxia promotes tumor cell growth. The study's main contribution lies in demonstrating how microfluidic oxygen control can replicate in vivo tumor conditions, allowing detailed observation of cell dynamics within physiological oxygen ranges. This technique provides a valuable platform for studying cancer progression and drug response in hypoxic microenvironments. For future work, the authors suggested enhancing the device's design to minimize oxygen diffusion errors and applying it to study metastasis, therapy resistance, and real-time tumor modeling under dynamic oxygen conditions [3].

Mishra, S., Yadav, D., and Sharma, S. (2023) proposed a DL based approach to improve accuracy and efficiency of breast cancer diagnosis from microscopic tissue images. Study introduces a patch-based CNN model that divides large histopathology images into smaller patches to better capture localized tissue features such as nuclei patterns and texture variations. The model was trained on benchmark datasets like BreakHis, enabling effective classification of benign and malignant tumors with high precision. This patch-wise analysis improves learning efficiency, reduces computational load, and enhances classification performance compared to traditional whole-image training. Main contribution of the work development of robust, scalable, and interpretable deep learning framework that can automatically extract discriminative features from histopathology slides without manual preprocessing. The authors reported superior accuracy and reliability in multi-class tissue classification tasks. For future work, they suggested integrating transfer learning and attention mechanisms to further enhance feature representation and applying the model to multi-modal diagnostic data for real-time clinical implementation [4].

Qureshi et al. (2024) presented comprehensive review of breast cancer detection.Study systematically explored evolution of breast cancer detection techniques from 1970 to 2023, emphasizing the transition from manual and image-processing methods to modern AI-driven deep learning frameworks. The authors followed the PRISMA guidelines to evaluate segmentation and classification methods for mammographic microcalcifications. They talked in details the three main aspects, namely, preprocessing, feature extraction, and classification, for image processing, ML, and DL techniques. The paper has been through the various techniques such as SVM, Decision Trees, CNNs, R-CNNs, and Transfer Learning models to show how deep learning architectures can beat traditional models in terms of both accuracy and speed. The paper's major contribution lies in providing a chronological and technical comparison of manual, ML, and DL-based mammography methods, highlighting large-scale datasets like DDSM, MIAS, and INBreast. The authors concluded that combining image preprocessing and advanced DL architectures significantly enhances early breast cancer detection. The future work emphasizes need for large annotated datasets, model interpretability, and explainable AI approaches to improve clinical applicability and trust in automated diagnostic systems [5].

Li, Y., Wu, J., and Wu, Q. (2019) In order to enhance multi class classification of breast cancer histopathology images, authors have presented a patch-based deep learning framework that not only extracts small (128×128) but also large (512×512) patches respectively. Initially, ResNet50 was used by them as a feature extractor and then K-means clustering was utilized along with CNN-based screening for selecting the most discriminative patches, thus making the model more accurate and less noisy due to the removal of irrelevant regions. The chosen features were combined with P-norm pooling and then classified by a Support Vector Machine (SVM), resulting in 95\% in total, thus leading to a performance that is better than the existing CNN-based methods. Primary contribution of this paper is novel multi-size patch extraction and discriminative patch selection method, which has a considerable effect on the diagnostic precision improvement. For future work, the authors suggested extending the framework to larger datasets and integrating more advanced deep architectures for real-time histopathological analysis in clinical environments [6].

This research paper presents a novel ML based method to detect breast cancer at early stage, utilizing Wisconsin Breast Cancer Diagnostic (WDBC) dataset. The authors have done an extensive data preprocessing work, which involves feature engineering, hypothesis testing, and Bonferroni-corrected feature selection. To handle the problem of class imbalance, they applied the SMOTE technique, and new features were created by the ensemble of seven classifiers. XGBoost classifier turned out to be best performer among fourteen models tested, with the performance metrics of 99.12\% accuracy, 0.9767 precision, 1.0 recall, 0.9861 specificity, and 0.9882 F1-score, thus, it surpassed the performance of the existing methods. Model also showed faster computation and higher reliability based on Cohen's Kappa. Overall, the proposed framework significantly improves breast cancer diagnosis accuracy and efficiency, highlighting XGBoost's effectiveness for medical classification tasks [7].

Strelcenia, E., and Prakoonwit, S. (2023) presented a study using GANs Breast Cancer Data" published in IEEE Access. The research focused on overcoming limited and imbalanced medical datasets in breast cancer diagnosis using a novel Kullback–Leibler Divergence Conditional GAN (K-CGAN). This GAN variant was trained to generate synthetic samples that balance benign and malignant cases in the Wisconsin Breast Cancer dataset. The authors compared K-CGAN with five GAN variants — LS-GAN, WGAN, NS-GAN, SDG-GAN, and traditional classifiers — and demonstrated superior classification performance across precision, recall, F1-score, and accuracy metrics. K-CGAN achieved high stability, with an F1-score above 0.99 and strong performance using classifiers such as XGBoost, Random Forest, and MLP. The study's key contribution lies in introducing a robust GAN-based data augmentation approach that enhances model generalization and classification reliability. For future work, the authors suggested extending the framework to larger and multimodal medical datasets and exploring hybrid deep learning–GAN combinations for real-world cancer diagnosis systems [8].

Aibe, N., Karasawa, et al.\ (2019) present findings from a large-scale web-based survey of 293 hospitals in Japan, aiming to understand trends and variations in postoperative radiotherapy (post-BCS RT) following breast-conserving surgery. The study found that 3D conformal radiotherapy (3DCRT) was used in 98\% of hospitals, with the field-in-field (FIF) technique applied in over one-third of cases to ensure homogeneous dose distribution. Most institutions followed conventional regimens of 50 Gy in 25 fractions for whole-breast irradiation and 10 Gy in 5 fractions for tumor bed boosts, while hypofractionated radiotherapy was implemented in 43of hospitals. The primary criteria for applying boost irradiation were surgical margin width (5 mm) and patient age (50 years). The study's major contribution lies in providing a nationwide overview of Japan's clinical radiotherapy practices, revealing both uniformity in conventional methods and emerging adoption of hypofractionated protocols.

Future work suggested tracking evolving trends, validating clinical outcomes, and aligning Japanese practices with updated international radiotherapy guidelines for improved patient care [9].

Aratake et al. (2023) proposed a microfluidic device to study how breast cancer cells behave under different oxygen concentration gradients, simulating the hypoxic tumor microenvironment. Their work combines numerical simulations with experimental validation to generate accurate oxygen profiles inside the device. Using 3D time-lapse imaging, they observed that cancer cells show maximum proliferation in regions with moderate hypoxia around 5 oxygen. The study found that proliferation, rather than migration, is the primary factor affecting cell distribution under oxygen gradients. Although the experimental results closely matched computational predictions, accuracy was limited at higher oxygen levels due to weak phosphorescence of sensing beads and slight oxygen leakage.

The research was also restricted to one cancer cell line and short-term observation. The authors highlight the need for better sensing materials and longer experimental durations. They also recommend testing multiple cell types and integrating molecular-level analysis. Overall, the study provides an important platform for understanding cancer behavior in controlled hypoxic conditions [10].

A. Problem Statement

Despite significant advancements in breast cancer screening and diagnosis, accurate interpretation of histopathology images remains a challenging and time-consuming task. Manual microscopic examination of Hematoxylin and Eosin (H&E) stained slides relies heavily on the expertise of pathologists and is subject to:

- Inter-observer variability
- Intra-observer inconsistency
- Fatigue-related errors
- Increasing workload due to rising cancer incidence

Existing computer-aided diagnosis (CAD) systems primarily focus on single-modality data (either histopathology or mammography) and often lack interpretability, limiting their acceptance in real clinical environments. Moreover, many deep learning models are dataset-specific and fail to generalize across magnification levels and imaging variations.

Therefore, there is a need for:

- A robust deep learning framework capable of accurately classifying breast cancer from histopathology images.
- A multimodal architecture that integrates microscopic and radiological features.
- An interpretable AI system that provides visual explanations to improve clinical trust.
- A scalable and generalized model that performs consistently across diverse datasets.

This research addresses these challenges by proposing an integrated, explainable, multimodal deep learning framework for reliable and clinically supportive breast cancer detection.

III. PROPOSED SYSTEM

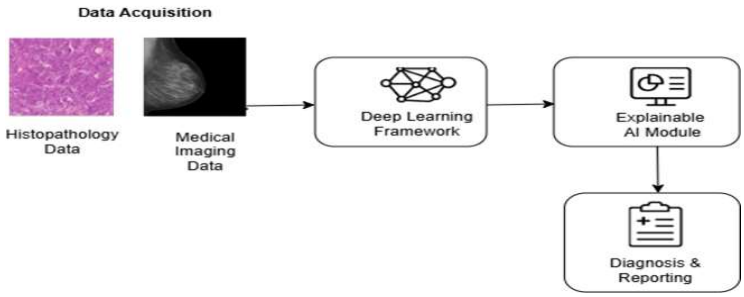

Figure 1. System Architecture

The proposed system architecture for DL based breast cancer detection is designed provide an automated, accurate, and interpretable diagnostic process. The architecture begins with the data acquisition stage, where histopathology and mammography images are collected to represent both microscopic and structural characteristics of breast tissue. These images are then passed through the data preprocessing stage, which involves noise removal, normalization, and segmentation to enhance image quality and isolate relevant regions for analysis. Next, feature extraction module employs a CNN to automatically learn important visual and spatial patterns that distinguish benign from malignant cells. The extracted features from different imaging modalities are then combined in the multimodal data fusion layer, which integrates information from various data sources to improve robustness and accuracy of model. The authors propose an aggregation mechanism which learns to combine noisy labels provided by the crowd and form high-quality training annotations instead of relying on expert-labeled data, which is constrained, time-intensive and expensive. AggNet architecture is a mitosis detector and label aggregator which is optimally trained together so as to learn robust features using imperfect datasets.

The next step is the classification module, which is based on deep learning or ensemble algorithms classify input data as benign or malignant. To improve transparency and credibility of the models, the architecture uses Explainable AI (XAI) module, which offers visual explanations, including heatmaps or activation maps, of which parts of the image were considered in making the choice. Lastly, processed results are sent to a Clinical Decision Support System (CDSS) which visualizes predictive diagnostic results and explanations which are offered to clinicians and allows them to make more informed and confident decisions. On the whole, the given architecture guarantees a holistic, end-to-end pipeline that integrates the multimodal data analysis, deep learning intelligence, and explainable outputs to guarantee the efficient detection of breast cancer and clinical guidance.

A. Dataset used

The BreakHis and BACH datasets provide histopathology images for microscopic-level analysis, while MIAS adds mammography images for radiological context. The WDBC dataset can be used for additional structured data integration and classification verification. These datasets combined can facilitate a multimodal learning method to allow the system to merge image and numerical data with better accuracy, resistance, and clinical significance in breast cancer diagnosis.

TABLE II. DATASETS USED AND THEIR CONSIDERATIONS

Dataset	Data Type	Considerations for This Study
BreakHis	Histopathology (microscopic) images	Provides benign and malignant breast tumor images at multiple magnification levels. Useful for patch-based CNN learning and feature extraction.
BACH (ICIAR)	Histopathology images	Contains high-resolution H&E-stained slides enabling cancer subtype classification. Suitable for model generalization and stain-normalization experiments.
MIAS	Mammography images	Adds radiological breast imaging for multimodal fusion. Helps evaluate cross-domain robustness and improves radiology-pathology correlation.
WDBC (Wisconsin Diagnostic Breast Cancer)	Structured numerical data	Provides clinical features (radius, texture, compactness, etc.). Useful for integrating tabular data with image features for improved diagnostic accuracy.

IV. MATHEMATICAL FORMULAS AND ALGORITHMS USED

A. Patch Extraction

Given an input histopathology image:

$$x \in R^{(H \times W \times C)} \quad (1)$$

The image is divided into N smaller patches:

$$p_k = Patch_k(x), \quad k = 1,2,...,N \quad (2)$$

B. Feature Extraction

Each patch is processed through a CNN feature extractor:

$$z_k = f_\theta(p_k) \quad (3)$$

where $z_k \in R^d$ represents the deep feature vector and θ denotes CNN parameters.

C. Patch-Level Classification

A fully connected classifier maps features to a score:

$$s_k = g_\varphi(z_k) \quad (4)$$

This score is converted to probability using sigmoid:

$$\hat{y}_k = 1 / (1 + e^{(-s_k)}) \quad (5)$$

where φ represents classifier parameters

D. Loss Function

Binary Cross-Entropy Loss:

$$L_{cls} = -(1/N) \Sigma [y_k \log(\hat{y}_k) + (1 - y_k) \log(1 - \hat{y}_k)] \quad (6)$$

E. Parameter Update

Parameters are optimized via gradient descent:

$$\Theta \leftarrow \Theta - \eta \nabla_\Theta L_{cls} \quad (7)$$

where $\Theta = \{\theta, \varphi\}$ and η is learning rate.

F. Image-Level Prediction

Final image classification using max pooling:

$$\hat{Y} = \max\_k \hat{y}\_k \quad (8)$$

Novelty of research

Novelty of this proposed system lies in integrated deep learning framework that combines multimodal medical data—including histopathology and mammography images—for more accurate and comprehensive breast cancer detection. Unlike traditional single-modality ordataset-specific models, this architecture fuses microscopic and radiological information to capture both cellular and structural tumor characteristics. Additionally, the inclusion of an Explainable AI (XAI) module enhances transparency by visually highlighting the regions influencing the model's decision, thereby improving clinical trust. The system's design further ensures real-time applicability and decision support for clinicians, making it a unique, interpretable, and reliable approach to early breast cancer diagnosis.

V. CONCLUSION

The reviewed studies highlight the significant advancements in automated breast cancer detection and analysis. DL based methods such as SE-Conformer, Patch-CNN, CBAM-VGGNet, AggNet, and Graph-TN have demonstrated potential to improve diagnostic accuracy and efficiency by reducing reliance on manual examination. Wearable-AI and clinical surveys give a supplementary information about the practice of early detection and treatment. Although these have been made, constraints to these innovations include reliance on high-quality datasets, computational aspects, semi-supervised learning, and inconsistencies in clinical application are still issues. Further studies are needed to enhance the ability to generalize the models to different populations, multi-modal modeling, and testing the approaches that have traditionally been used under varied clinical conditions in order to maximize the applicability of the model in practice.

REFERENCES

[1] L. S. Nair, K. R. Amarnath, and J. J. Nair, "SE-Conformer framework of malignancy detection on histopathology images," IEEE Access, vol. 13, pp. 100105–100118, 2025, doi: 10.1109/ACCESS.2025.3576825.

[2] D. N. Elsheakh, O. M. Fahmy, M. Farouk, K. Ezzat, and A. R. Eldamak, "Early breast cancer detection using wearable flexible sensors and artificial intelligence," IEEE Access, vol. 12, pp. 48511–48529, 2024, doi: 10.1109/ACCESS.2024.3380453.

[3] S. Aratake, N. Kawahara, and K. Funamoto, "Behavior of breast cancer cells under oxygen concentration gradients in a microfluidic device," Proc. IEEE EMBC, pp. 1–6, 2023, doi: 10.1109/EMBC40787.2023.10340327.

[4] S. Mishra, D. Yadav, and S. Sharma, "Breast cancer classification from histopathological images using patch-based deep learning modeling," IEEE Access, vol. 11, pp. 85742–85755, 2023, doi: 10.1109/ACCESS.2023.3289456.

[5] S. A. Qureshi et al., "Breast cancer detection with mammography: Image processing to deep learning," IEEE Access, vol. 13, pp. 60776–60792, 2024, doi: 10.1109/ACCESS.2024.3523745.

[6] Y. Li, J. Wu, and Q. Wu, "Classification of breast cancer histology image based on multi-size and discriminative patch," IEEE Access, vol. 7, pp. 21400–21408, 2019, doi: 10.1109/ACCESS.2019.2898044.

[7] M. Rai et al., "Improving early-stage breast cancer detection using advanced data analysis algorithms," IEEE Access, vol. 12, pp. 78432–78445, 2024.

[8] E. Strelcenia and S. Prakoonwit, "Improving cancer detection classification with GANs on breast cancer data," IEEE Access, vol. 11, pp. 71594–71614, 2023.

[9] N. Aibe et al., "Nationwide survey on Japanese clinical practice in breast-conserving radiotherapy," J. Radiation Research, vol. 60, no. 1, pp. 142–149, 2019, doi: 10.1093/jrr/rry095.

[10] T. L. Dinger et al., "Assessment of response to primary endocrine therapy in elderly breast cancer patients," Breast Cancer Research and Treatment, accepted Aug. 2025.