International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

AI-Driven Histopathological Image Analysis for Genetic Evaluation of Breast Cancer

Ayush Vanjare[1], Mohamedalmodather[2], Srivaramangai R[3]
[1,2,3]*Department of information Technology, University of Mumbai, India*

***Abstract*-- Breast cancer is a predominant worldwide challenge that continues to necessitate an evolving approach towards diagnosis and prognostication. With regards to diagnostics, classical paradigms of histopathology supported by limited biomarker analysis are increasingly being questioned by the biological complexities of cancer and its inherent molecular heterogeneity. In this review, we summarize the latest advancements of AI, ML, and DL in breast cancer genomics, transcriptomics and computational pathology. A variety of AI approaches are reviewed which span from feature-optimized traditional ML models, to deep convolutional neural networks, transformer-based embeddings, and task-oriented computational pathology paradigms. Many of the studies discussed in this review show that contemporary AI models can attain greater than 98% diagnostic accuracies and, in some circumstances, can directly predict prognostically significant molecular features such as BRCA1/2 mutations from standard histological images. The recent emphasis on XAI is highlighted as a fundamental prerequisite in ensuring trust, clinical utility and regulatory approval of AI-driven approaches in breast cancer diagnostics and beyond.**

Keywords **– artificial intelligence, Breast cancer, computational Pathology, Genomics, Transcriptomics, explainable AI, Precision Oncology.**

I. Introduction

Globally, breast cancer is the most frequent female cancer and a major cause of cancer-related death. It is documented as the incidence of breast cancer has been rising constantly both in the high-income and the low- to middle-income countries due to several factors like changes in demography, lifestyles, better methods of diagnosis etc. [1] [2]. Nevertheless the results vary greatly because of the heterogeneous biology and molecular basis of breast cancer despite advances in early diagnosis and therapy. The most distinctive feature of breast cancer is its substantial heterogeneity, histologically, molecularly and clinically.Tumors that exhibit similar morphological features on routine histopathology could in fact have completely different genetic mutations, gene expression profiles, responses to therapy and outcome.Breast cancer is not a single disease but consist of several molecular subtypes, including Luminal A,Luminal B,HER2-enriched, and Triple-Negative Breast Cancer(TNBC).Each subtype has distinct biological characteristics and responds differently to available treatments [3] [4].

This variation makes it challenging to develop a standardized approach for diagnosis and treatment planning, highlighting the need for more personalized and accurate clinical strategies.

Genetic predisposition drives breast cancer initiation and progression. Overwhelmingly most breast cancers develop sporadically; however, 5-10% of breast cancers develop familial breast cancer through the inherited transmission of germline mutations within specific genes which are known to significantly increase the risk of developing breast cancer [5]. In particular BRCA1 and BRCA2 are the best characterized high-penetrance genes implicated within breast and ovarian cancer syndrome; carrying a harmful BRCA1/2 mutation increases lifetime risk of breast cancer significantly [6]. The risk genes for breast cancer TP53, PALB2, CHEK2, ATM and PTEN are just another component adding further complexity; leading to a polygenic basis for breast cancer risk [7]. Although these mutations can be identified through advances in next generation sequencing and transcriptomics there is still a significant challenge to incorporate this data into every-day clinical decision making. However, the standard clinical pathways involving histopathology and immunohistochemistry is still subject to inter-observer variations and subjective interpretation and have limitations on the scalability. Due to the aforesaid limitations, there has been growing interest in computational tools that could leverage the morphology, molecule and clinical information. AI presents itself as a powerful technology in this domain; it has the ability to identify non-linear relationships and intricate associations in huge biomedical datasets that are otherwise challenging to establish using traditional statistics approaches [8] [9]. In recent reports, it has been shown that AI based methods could even directly predict the molecular features directly from histological images which could challenge the rigid segregation of tissue morphology and genomic profile [10] [11] [12].

II. Literature Review

A. Molecularly Guided Therapeutic Strategies

The understanding of molecular basis of breast cancer has revolutionised therapeutic choices for breast cancer patients. Clinical trials have yielded evidence that specifically targeting genetic abnormalities is beneficial for patients.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

Andr et al. Demonstrated that treatment of PIK3CA-mutated ER-positive metastatic breast cancer with the PI3K inhibitor alpelisib increased progression free survival [13]. Tutt et al. Also found that in high risk ER-negative breast cancer with germline BRCA1/2 mutations, adding olaparib as adjuvant treatment decreased the risk of recurrence [14]. These studies highlight the clinical advantage of meticulous defining molecular features in guiding individual cancer treatments.

B. Machine Learning in Breast Cancer Diagnosis

Machine learning applications were first applied to the diagnosis of breast cancer using the structured clinical and gene-expression data. Houfani et al applied various ML algorithms to the Wisconsin Diagnostic Breast Cancer dataset, and reported classification accuracies of 98% by multilayer perceptron and logistic regression models [15]. Wu and Hicks were the first to expand ML based classification to transcriptomic data and used RNA-Seq data on TNBC samples from TCGA, achieving 90% classification of triple-negative versus non-TNBC samples using RNA-Seq features [16]. The use of feature selection techniques such as Neighborhood Component Analysis and Minimum Redundancy Maximum Relevance boosted classification accuracy and interpretability [17] [18].

C. Deep Learning and Computational Pathology

Image-based breast cancer diagnosis has benefited greatly from the advent of deep learning techniques. Trained on whole-slide histopathology images, convolutional neural networks can automatically learn discriminatory features without the need for manual annotation. It has been shown that a ResNet based CNN trained on H&E stained whole slide images was able to classify germline BRCA status, attaining an AUC of 0.906 on high-grade tumors [10]. Using tumor segmentation and contrastive learning, Bourgade et al. Significantly enhanced performance further [11]. The task-oriented computational pathology platforms that has been developed recently achieve near-perfect classification performance with AUC of 0.999. [12] .

D. Explainable Artificial Intelligence

Although high predictive accuracy was achieved, adoption in clinical practice is limited because many AI models operate as black-boxes. To mitigate this problem, explainable AI models are needed that can offer interpretable information regarding a prediction. Kallah-Dagadu et al. Developed interpretable ML models that showed more than 99% accuracy using SHAP values and partial dependence plots without compromising the interpretability of the model [18]. Explainable AI is now perceived as crucial for ethical implementation and regulatory compliance [19].

III. Methodology / Proposed System

We present here an AI-based approach aiming to predict genes and molecules based on histopathological data for the characterization of breast cancer. As opposed to current clinical practice based on an initial microscopic examination of tissue morphology (and a subsequent molecular analysis), this framework directly infers genetic information from the images by combining the field of computational pathology with machine learning. The purpose of this approach is to take advantage of the morphological features observed in breast cancer images associated with the molecular modifications in order to achieve a non-invasive and less costly determination of genetic profile. The breast cancer histopathology images for the present investigation were obtained from a dataset that has been constructed in order to perform the evaluation of genetics. Such datasets are beginning to gain attention for their ability to encode dormant morphological fingerprints in relation to genetic and transcriptomic alterations such as BRCA1/2 mutation status, molecular subtype and pathway deregulation. Several research works have shown that models that utilize deep learning architectures are capable of extracting these subtle correlations in images stained by hematoxylin and eosin, thereby inspiring the use of image based learning for inferring molecular signatures [10] [11] [12]. Preprocessing histopathological images is carried out before training a model to standardize and improve robustness. The steps include normalization of resolution, color normalization, to correct for color deviation between stainings and, to some extent, artifacts without clinically meaningful features. Whole-slide images are divided into smaller images or patches, a common practice in computational pathology [10]. Deep networks can analyze high resolution information of tissue, while keeping local contextual structure by dividing images into patches [10]. We performed feature extraction using deep convolutional neural networks that learned hierarchical features from image data without manual feature engineering. Residual architectures such as were appropriate because they were able to learn high level spatial patterns and also avoid the problems of vanishing gradients in deep networks. The learned features represented cell structure, tissue structure, and stromal context that would be very difficult to measure manually, and were known to be correlated with molecular phenotypes [11] [12]. Features obtained from this procedure were then applied to genetic prediction problem which consists of breast cancer classification and molecular characterization. There are two modes for binary or multi-class classification depending on the goal: cancer vs normal classification, molecular subtypes discrimination or mutation positive/negative decision.

Features are trained by supervised learning and labels come from the related genetic/molecular annotations. Evaluation is carried out with well-known metrics like accuracy, sensitivity, specificity, AUC of the ROC curve in order to guarantee clinical significance and comparability with other papers [15] [20] [21]. To enhance clinical interpretability, the proposed system incorporates explainable artificial intelligence techniques that highlight image regions contributing most significantly to model predictions. Visualization mechanisms such as attention maps and saliency-based explanations allow clinicians to relate algorithmic decisions to recognizable histopathological structures. The inclusion of explainability addresses one of the primary barriers to clinical adoption of deep learning models in medical diagnostics by improving transparency and trust [19] [18].

Overall, the proposed system provides an image-based framework for breast cancer evaluation with a focus on predicting genetic characteristics. It helps bridge the gap between traditional histopathology and molecular diagnostics by using artificial intelligence to identify meaningful patterns in tissue images that may be associated with genetic changes. This approach has the potential to support faster and less invasive clinical decision-making, contribute to more personalized treatment strategies, and advance the shift from conventional pathology toward predictive, data-driven cancer care.

IV. Discussion

The discoveries and algorithms that have been explored and applied in this research illustrate the increasing aptitude of artificial intelligence to narrow the centuries-old rift between histopathological interpretation and molecular profiling of breast cancer. The system described herein, in that it uses an image-based dataset with a genetic intention, follows a general movement away from traditional and time-consuming/expensive/limited techniques like next-generation sequencing, toward a non-invasive, data-driven inference of molecular changes within cancer. There is growing evidence that tissue morphology captures latent genomic and transcriptomic profiles of tumors, serving as a proxy for molecular information. Deep learning models are capable of extracting subtle morphologic features that require manual quantification to be described, such as different levels of cellular and nuclear atypia, and stromal architecture. To date, these models have been shown to be able to predict therapeutically important genetic mutations, such as BRCA1/2 status from only Hematoxylin-and-Eosin-stained pathology slides [10] [11] [12], and it appears that the hypothesis underlying the present investigation is also grounded in such genotype-phenotype correlations.

The advantages of deep learning-based computational pathology over traditional machine learning methods based on hand-crafted features and structured molecular data include its high scalability and flexibility. While the latter method performs reasonably well on selected gene expression datasets, classifying with relatively high accuracy when combined with strong feature selection [15] [17] [18], it can be extended only to situations in which there are measurements for each specific molecule. This limitation will restrict their utility in cases when genetic testing is not accessible or desirable. Instead, we investigated the computational approach described here that uses images as input to provide molecular information and thus the benefit of this new approach can be extendable to broader population in oncology. These benefits should however not obscure a few important considerations. A major drawback pertains to generalizability across different institutions and population. Image acquisition characteristics like differences in tissue processing, staining, scanner hardware can all contribute to performance variation of the models. Most of the aforementioned works used retrospective, multi-center data from few sites and a concern about underlying biases and weaker performance in the new setting arises [9] [19]. More multi-institutional and diverse datasets are essential and standardization of the preprocessing pipelines is needed to avoid such shortcomings. Yet another critical aspect is model interpretability. The predictive performance is undoubtedly paramount; however, acceptance in the clinical workflow can only be achieved if a system can be reliably understood and trusted. Black-box models that output predictions without justification may not be fully trusted by both clinicians and regulatory agencies. Thus, the utilization of explainable AI methods is one of the key steps taken by the presented system to address this shortcoming. Visualization of important regions in the image and attention mechanism help doctors make connections between algorithmic predictions and identifiable histopathological features while maintaining good accuracy [19] [18]. Another significant implication of embedding genetic intent in image-based analysis has to do with the clinical workflow. Such a system could be used as a triage mechanism if the results can be validated widely, enabling clinicians to know which patients are more likely to carry certain genetic alterations and in turn warrant a confirmatory molecular test or targeted therapy. Such an system would also likely help with efficient allocation of resources and expedite clinical decision making, particularly in low-resource settings in which broad genome profiling may not be readily available. To conclude, the findings presented reinforce the role of AI as a tool that complements and does not supplant the expertise of the pathologist or geneticist.

AI-powered tools that fuse the power of computation with the subtleties of clinical knowledge offer an opportunity to achieve more consistent, objective, and individualized care of patients with breast cancer. Further advancements will require continued interdisciplinary collaboration and careful validation as well as attention to ethical and regulatory issues.

V. Conclusion And Future Scope

In this work we highlight the impact that artificial intelligence can have to transform the diagnosis of breast cancer from morphological features to genetically driven decisions (precision medicine). We illustrate the capability of using histopathological images to genetically and molecularly classify breast cancer tumors in the presented framework, and also to extract tumor molecular features using image information without need of expensive and separated molecular testing procedures, which are common requirements of current work-flow for such information to be gained. Computational pathology, therefore combined with AI, holds promise for the future diagnosis of breast cancer. The increasing body of evidence described and considered here supports the claim that DL-based models can infer latent genotype-phenotype associations underlying tissue morphology. The fact that research utilizing these models can achieve a high diagnostic accuracy and reliably predict molecular profiles confirms that inference from image data can be highly effective in many scenarios, especially regarding clinically relevant changes such as mutations in BRCA1/2 genes and molecular subtypes [15] [10] [11] [20] [12]. This reiterates the concept that histopathology images are not mere artifacts but, instead, contain substantial biological data that can be systematically unlocked with smart computing models. In addition to performance, another key implication from this work is the value of interpretability and clinical trust. Using an XAI component for clinical applications solves a major issue with adoption by making model outputs comprehensible. The need for interpretability in AI systems, which can be a technical barrier to adoption, has also been highlighted in several recent papers as an ethical and regulatory necessity [19] [18]. A system that offers both high performance and an ability to interpret its outputs can contribute to clinical acceptance and approval. The field continues to move forward on a variety of frontiers for future research. Integration of multimodal data, like histopathology, genomics, transcriptomics, radiological imaging, and clinical records, is one path with the potential to enhance both predictive reliability and clinical meaningfulness. These frameworks are thought to capture a more complete view of tumor biology that can't be learned from individual modes alone [12]. Such an undertaking, however, demands progress in data standardization, harmonization and computational efficiency.

Secondly, future work should address the validation of models on large, multi-institutional datasets. Currently, most studies employ retrospective data obtained from small sample sets, which may be prone to bias and not generalizable across heterogeneous patient groups [9] [19]. By combining patient data from geographically and demographically disparate institutions, fair, robust and clinically applicable AI diagnostics can be developed, the realization of which is facilitated by new trends such as federated learning and privacy-preserving AI. In summary, the application of artificial intelligence for histopathological images to study genetics is a promising, dynamic approach in breast cancer management that has evolved from research into potential clinical use as these methods develop and extensive validation is conducted. Integration into routine practice will be a multifaceted process of continued interdisciplinary communication, critical validation and an ongoing commitment to transparent and ethical implementation. In conclusion, using intelligent computations to analyze breast cancer histopathology can have an impact on earlier detection and better treatment selection and patient management in the context of personalized medicine.

REFERENCES

[1] World Health Organization, "Assessment of Breast Cancer Control Capacities in the WHO African Region in 2022," World Health Organization, Geneva, 2024.

[2] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 71, no. 3, p. 209–249, 2021.

[3] C. M. Perou, T. Sørlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees and J. R. Pollack, "Molecular portraits of human breast tumours," Nature, vol. 406, no. 6797, p. 747–752, 2000.

[4] R. Lupat, R. Perera, S. Loi and J. Li, "Moanna: Multi-omics Autoencoder-Based Neural Network Algorithm for Predicting Breast Cancer Subtypes," IEEE Access, vol. 11, p. 10912–10924, 2023.

[5] N. Mavaddat, A. C. Antoniou, D. F. Easton and M. Garcia-Closas, "Genetic susceptibility to breast cancer," Molecular Oncology, vol. 4, no. 3, pp. 174-191, 2010.

[6] K. B. Kuchenbaecker, J. L. Hopper, D. R. Barnes, K.-A. Phillips, T. M. Mooij, M.-J. Roos-Blom and S. Jervis, "Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers," JAMA, vol. 317, no. 23, pp. 2402-2416, 2017.

[7] F. J. Couch, H. Shimelis, C. Hu, S. N. Hart, E. C. Polley, J. Na and E. Hallberg, "Associations between cancer predisposition testing panel genes and breast cancer," JAMA Oncology, vol. 3, no. 9, pp. 1190-1196, 2017.

[8] R. Dias and A. Torkamani, "Artificial intelligence in clinical and genomic diagnostics," Genome Medicine, vol. 11, no. 1, p. 70, 2019.

[9] S. Laso, D. Flores-Martin, J. L. Herrera, J. Galan-Jimenez and J. Berrocal, "Identification and visualization of a patient's medical record via mobile devices without an internet connection," Electronics, vol. 12, no. 1, p. 75, 2022.

[10] X. Wang, C. Zou, Y. Zhang, X. Li, C. Wang, F. Ke and J. Chen, "Prediction of BRCA gene mutation in breast cancer based on deep learning and histopathology images," Frontiers in Genetics, vol. 12, p. 661109, 2021.

[11] R. Bourgade, N. Rabilloud, T. Perennec, T. Pécot, C. Garrec, A. F. Guedon and C. Delnatte, "Deep learning for detecting BRCA mutations in high-grade ovarian cancer based on an innovative tumor segmentation method from whole-slide images," Modern Pathology, vol. 36, no. 11, p. 100304, 2023.

[12] C. Frascarelli, K. Venetis, A. Marra, A. Concardi, M. D'Ercole, E. Mangione and M. Negrelli, "Computational pathology in breast cancer: optimizing molecular prediction through task-oriented AI models," NPJ Breast Cancer, vol. 11, no. 1, p. 141, 2025.

[13] F. André, E. Ciruelos, G. Rubovszky, M. Campone, S. Loibl, H. S. Rugo and H. Iwata, "Alpelisib for PIK3CA-mutated, hormone receptor-positive advanced breast cancer," New England Journal of Medicine, vol. 380, no. 20, pp. 1929-1940, 2019.

[14] A. N. J. Tutt, J. E. Garber, B. Kaufman, G. Viale, D. Fumagalli, P. Rastogi and R. D. Gelber, "Adjuvant olaparib for patients with BRCA1- or BRCA2-mutated breast cancer," New England Journal of Medicine, vol. 384, no. 25, pp. 2394-2405, 2021.

[15] D. Houfani, S. Slatnia, O. Kazar, N. Zerhouni, H. Saouli and I. Remadna, "Breast cancer classification using machine learning techniques: a comparative study," Medical Technologies Journal, vol. 4, no. 2, pp. 535-544, 2020.

[16] J. Wu and C. Hicks, "Breast cancer type classification using machine learning," Journal of Personalized Medicine, vol. 11, no. 2, p. 61, 2021.

[17] A. Abidalkareem, A. K. Ibrahim, M. Abd, O. Rehman and H. Zhuang, "Identification of gene expression in different stages of breast cancer with machine learning," Cancers, vol. 16, no. 10, p. 1864, 2024.

[18] G. Kallah-Dagadu, M. Mohammed, J. B. Nasejje, N. N. Mchunu, H. S. Twabi, J. M. Batidzirai, G. C. Singini, P. Nevhungoni and I. Maposa, "Breast cancer prediction based on gene expression data using interpretable machine learning techniques," Scientific Reports, vol. 15, no. 1, p. 7594, 2025.

[19] N. Ghaffar Nia, E. Kaplanoglu and A. Nasab, "Evaluation of artificial intelligence techniques in disease diagnosis and prediction," Discover Artificial Intelligence, vol. 3, no. 1, p. 5, 2023.

[20] S. Muthamilselvan and A. Palaniappan, "BRCADX: Precise identification of breast cancer from expression data using a minimal set of features," Frontiers in Bioinformatics, vol. 3, p. 1103493, 2023.

[21] S. Thirumalaisamy, K. Thangavilou, H. Rajadurai, O. Saidani, N. Alturki, S. K. Mathivanan, P. Jayagopal and S. Gochhait, "Breast Cancer Classification Using Synthesized Deep Learning Model with Metaheuristic Optimization Algorithm," Diagnostics, vol. 13, no. 18, 2023.