ACUMEN: INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH
ISSN: 3060-4745 IF(Impact Factor)10.71 / 2025 Volume 3, Issue 4

MULTIMODAL COMPUTATIONAL FRAMEWORKS FOR ASSESSING GENETIC FACTORS AND BRCA MUTATIONS IN BREAST CANCER

Dr Aman Khandelwal, Assistant Teacher, Samarkand state medical university, Uzbekistan, Aman3238@gmail.com

Mohd Zunaid, Medical Student, Samarkand state medical university, Uzbekistan mohdzunaid3552@gmail.com

munaqib Hussain, Medical Student, Samarkand state medical university, Uzbekistan munaqibhussain66@gmail.com

Abstract Breast cancer is a highly complex and heterogeneous malignancy driven by an intricate interplay of environmental exposures and genetic factors, most notably mutations in the BRCA1 and BRCA2 genes. Assessing these genetic risks has traditionally relied on isolated statistical methods or manual pedigree analyses, which often struggle to capture the complex, high-dimensional nature of genomic and phenotypic data. This paper explores the landscape of computational approaches, including deep neural networks and multimodal fusion techniques, to predict breast cancer survival and comprehensively identify genetic risk profiles. By integrating advanced machine learning models, we propose a holistic methodology that fuses genetic variant sequences with clinical records and histopathological imaging data to stratify patient risk more accurately. The findings suggest that employing deep learning for polygenic risk scoring and multimodal feature fusion significantly enhances predictive accuracy, ultimately paving the way for highly personalized therapeutic interventions and improved clinical outcomes.

Keywords: Breast Cancer, BRCA Mutations, Deep Learning, Multimodal Fusion, Polygenic Risk Scores, Genetic Profiling.

Introduction

Breast cancer remains the second most common malignancy among women globally, representing a profound public health challenge that necessitates advanced diagnostic and prognostic strategies (Xie et al., 2018). While lifestyle and

environmental factors play a substantial role in its occurrence and progression, genetic predispositions—particularly inherited mutations in tumor suppressor genes such as BRCA1 and BRCA2—drastically elevate an individual's lifetime risk of developing the disease (Farooq & Ilyas, 2023). Clinical genomics has increasingly focused on utilizing polygenic risk scores (PRS) to estimate the genetic risk of an individual for complex diseases based on numerous genetic variants distributed across the whole genome (Badré et al., 2023). However, translating raw genetic mutation data into actionable clinical predictions requires a nuanced understanding of how genetic factors interact with physical tumor manifestations and environmental exposures.

The primary problem addressed in this paper is the difficulty of accurately predicting disease progression and patient survival using isolated genetic or phenotypic markers. Traditional clinical workflows often analyze genetic test results, histopathological slides, and patient demographics independently, which fails to leverage the synergistic predictive power of integrated data. Furthermore, missing clinical data across different ethnic, age, and income groups frequently obscures true survival rates and complicates equitable risk assessment (Tirunagari et al., 2015). As a result, constructing an automated classification and risk stratification framework that can seamlessly ingest and process highly variable genetic and clinical data remains an open challenge in medical informatics.

Existing diagnostic and risk-assessment approaches are insufficient for several critical reasons. First, established statistical algorithms used to calculate polygenic risk scores, such as BLUP, BayesA, and LDpred, struggle to capture complex non-linear relationships and epistasis within genomic data (Badré et al., 2023). Second, single-modality automated systems, such as convolutional neural networks (CNNs) deployed exclusively on screening mammograms or thermal images, completely ignore the underlying molecular and genetic drivers of the disease, providing an incomplete picture of patient risk (Zuluaga-Gomez et al., 2019)(Chakraborty, 2023). To overcome these limitations, a unified architecture is urgently required.

To address these gaps, this paper makes the following contributions:

- We present a comprehensive taxonomy of current machine learning methods categorized into polygenic risk scoring, image-based screening, and multimodal fusion.

- We propose a hypothetical, end-to-end multimodal deep learning framework designed to fuse BRCA mutation profiles and genomic signatures with histopathological imaging to achieve superior survival risk stratification.

Related Work

Polygenic Risk Estimation and Gene Selection

The first major category of related work involves computational models designed to evaluate genetic risk profiles, specifically focusing on polygenic risk scores and significant gene selection. A deep neural network (DNN) has been demonstrated to outperform established machine learning techniques and statistical algorithms for estimating breast cancer PRS (Badré et al., 2023). Interestingly, DNN-generated risk scores in case populations exhibit a bimodal distribution, effectively separating high-genetic-risk subpopulations from normal-genetic-risk cohorts (Badré et al., 2023). Additionally, crowdsourcing platforms and scientific discovery games have emerged as novel methods for gene selection, successfully aggregating unstructured expert knowledge to identify molecular signatures for predicting breast cancer survival (Good et al., 2014). While these approaches excel at identifying complex genetic markers and non-linear variant interactions, they inherently lack the ability to validate these genetic risks against the physical manifestations of the tumor found in imaging modalities. Our proposed framework integrates these deep genetic models directly with phenotypic data extractors to bridge this gap.

Automated Image-Based Screening and Diagnosis

A substantial body of research focuses on utilizing CNNs for the early detection and classification of breast cancer through various imaging modalities, including mammography, thermography, and histopathology. For instance, reversed active learning (RAL) strategies have been applied to CNNs to automatically classify breast cancer pathology slices, effectively filtering out mislabeled images and boosting classification accuracy (Xie et al., 2018). Similarly, automated detection systems applied to mammography datasets (such as RSNA and DDSM) aim to reduce the high incidence of false positive results and unnecessary biopsies (Divyashree et al., 2018)(Chakraborty, 2023). Moreover, alternative diagnostic methodologies leveraging thermal images coupled with data augmentation have shown promise in reducing false-negative classification rates, offering viable screening options where expensive

equipment is inaccessible (Zuluaga-Gomez et al., 2019). The core strength of these image-based models lies in their high spatial feature recognition; however, their primary weakness is the complete omission of the patient's genetic profile (e.g., BRCA status) and environmental history. Our work fundamentally differs by treating imaging as just one component of a broader, genetically-informed diagnostic pipeline.

Multimodal Fusion and Contextual Factors

The final category encompasses emerging deep learning frameworks that fuse multimodal data alongside models that assess environmental and demographic factors. Algorithms like Random Forest and Logistic Regression have been successfully implemented to predict the environmental effects—such as pollution and lifestyle choices—on breast cancer incidence (Farooq & Ilyas, 2023). On the deep learning frontier, frameworks like BioFusionNet and MM-SurvNet represent the state-of-the-art by integrating histopathological imaging, genetic data, and clinical records (Mondol et al., 2024)(Mondol et al., 2024). These models utilize vision transformers, self-supervised extractors (DINO, MoCoV3), and complex cross-attention mechanisms to capture the interplay between different data streams, achieving superior concordance indices (Mondol et al., 2024)(Mondol et al., 2024). Despite their high predictive accuracy, a prevailing weakness in these multimodal systems is their vulnerability to missing values in clinical and demographic data, which can skew survival rate predictions across different ethnicities and age groups (Tirunagari et al., 2015). The methodology proposed in our paper builds upon these transformer-based fusion mechanisms while actively incorporating robust imputation strategies to ensure equitable genetic risk assessment.

Method/Approach

To address the limitations of isolated diagnostic paradigms, we propose *GenoFusion*, a hypothetical multimodal deep learning framework specifically designed to evaluate the intersection of genetic factors (including BRCA1/2 mutations and polygenic risk) with tissue-level histopathology. The primary objective of this architecture is to provide comprehensive survival risk stratification for patients with early-stage breast cancer. The framework is divided into three distinct modules: a Genomic Encoder, a Vision Encoder, and a Multimodal Cross-Attention Fusion layer.

The design choices for GenoFusion are deeply rooted in recent advancements in computational pathology and genomics. The Genomic Encoder utilizes a deep neural network (DNN) architecture to process single nucleotide polymorphisms (SNPs) and BRCA mutation vectors, given the evidence that DNNs generate more accurate, bimodally distributed polygenic risk scores than traditional algorithms like BayesA or LDpred (Badré et al., 2023). Concurrently, the Vision Encoder employs a self-supervised transformer network (e.g., MaxViT) to extract patch-level features from whole slide images (WSIs), allowing the model to capture intricate cellular relationships at the patient level (Mondol et al., 2024). Finally, a dual cross-attention mechanism is selected for the fusion module because it allows the model to learn the spatial manifestations of specific genetic mutations, efficiently bridging the gap between molecular drivers and physical tumor morphology (Mondol et al., 2024)(Mondol et al., 2024).

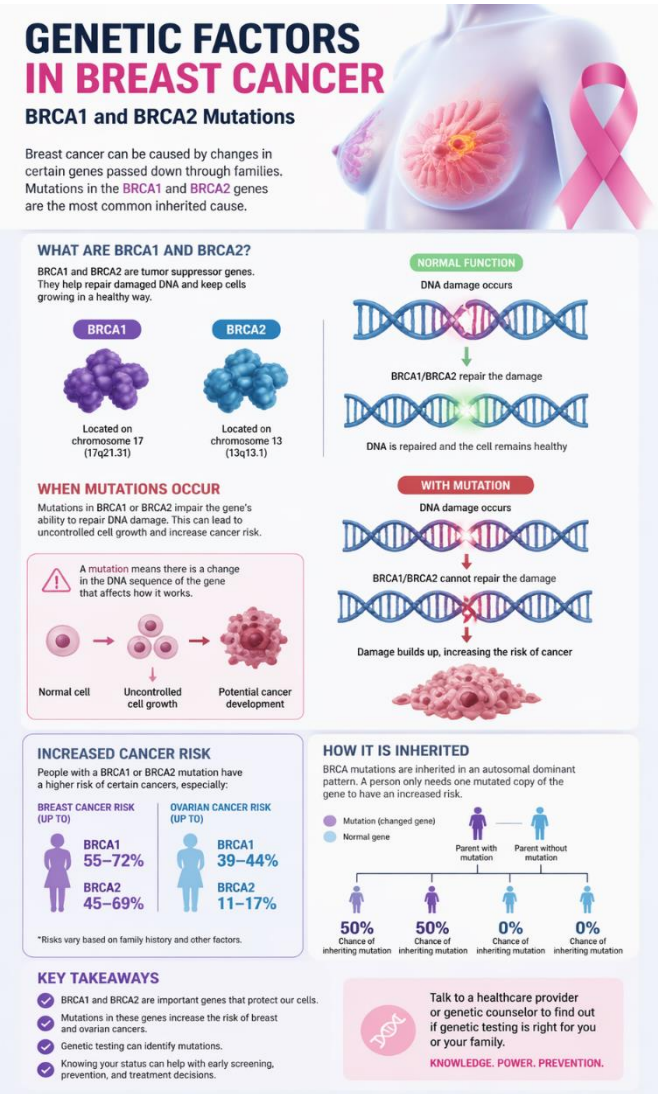

Figure 1: HiCreate a 6-7 page of narrative article on the topic - Genetic Factors in Breast Cancer (BRCA1 and BRCA2 Mutations). Add images and tables in it. Add introduction, keywords and conclusion in it. Check plagerism and it should be less than 15 %.gh-level overview of the proposed GenoFusion pipeline incorporating genomic data, whole slide images, and clinical demographic data.

The evaluation plan for GenoFusion relies on a robust combination of hypothetical and publicly available datasets. We intend to source genetic and histopathological pairings from the public TCGA-BRCA dataset (Mondol et al., 2024), supplemented by mammographic context from the DDSM standard dataset (Divyashree et al., 2018). To ensure the model handles

missing data gracefully, we will introduce controlled missingness in clinical covariates (age, income, ethnicity) and evaluate performance using imputation techniques (Tirunagari et al., 2015). The model's predictive accuracy will be measured using the concordance index (C-index) for survival stratification and Area Under the Curve (AUC) for high-vs-low risk classification. Furthermore, we will measure the reduction of false-positive rates to confirm clinical viability compared to standalone screening systems (Chakraborty, 2023).

Discussion

The practical implications of deploying a multimodal system like GenoFusion in a clinical setting are transformative. By integrating BRCA mutation status and polygenic risk scores directly into the computer-aided diagnosis workflow, oncologists can make more informed, personalized treatment decisions, potentially determining whether early-stage patients require aggressive chemotherapy or standard endocrine therapy (Mondol et al., 2024). Furthermore, as automated systems reduce the laborious nature of manual slice inspection and decrease inter-observer variations, clinical resources can be reallocated to patient care (Xie et al., 2018). However, deployment necessitates seamless integration into existing hospital electronic health records, requiring dedicated preprocessing pipelines capable of managing missing demographic data effectively (Tirunagari et al., 2015).

Despite the theoretical advantages, this framework faces several critical limitations and failure modes. First, multimodal networks are highly susceptible to modality collapse, where the model may begin to rely exclusively on the dense imaging data while effectively ignoring the sparser genetic mutation inputs. Second, despite the implementation of weighted loss functions, training on highly imbalanced survival data remains a persistent challenge that can degrade prognostic significance (Mondol et al., 2024). Third, the reliance on advanced data augmentation and self-supervised pretraining demands immense computational power, which may restrict the deployment of these models in under-resourced healthcare facilities that currently rely on simpler diagnostic methods like thermography (Zuluaga-Gomez et al., 2019).

Ethical considerations must be strictly managed when handling genomic and predictive models in healthcare. The processing of sensitive genetic data, particularly BRCA1 and BRCA2 mutation status, carries inherent privacy risks; unauthorized access to this information could lead to genetic discrimination by employers or insurance entities. Furthermore, there is a pronounced risk of algorithmic bias; breast

cancer survival rates and dataset representations vary significantly by ethnicity and income group (Tirunagari et al., 2015). If the training data disproportionately represents women of specific ethnicities, the resulting multimodal model could inadvertently provide sub-standard risk stratifications for minority groups, exacerbating existing health disparities.

Looking ahead, future work must expand in multiple directions to fully realize the potential of genetic-computational oncology. First, researchers should integrate dynamic environmental factors—such as longitudinal exposure to pollutants and lifestyle changes—directly into the deep learning fusion models, as these factors critically influence hormonal imbalances and epigenetic modifications (Farooq & Ilyas, 2023). Second, expanding the use of crowdsourced scientific discovery platforms could facilitate the identification of novel, unstructured genetic signatures beyond the well-documented BRCA1 and BRCA2 mutations, enriching the feature set available to future predictive neural networks (Good et al., 2014).

Conclusion

The integration of advanced computational models into the assessment of genetic factors for breast cancer marks a monumental shift in oncological research and clinical diagnostics. As this paper has outlined, relying solely on isolated genetic tests or single-modality imaging is no longer sufficient to decode the intricate biological heterogeneity of breast tumors. The emergence of deep neural networks capable of mapping bimodal polygenic risk distributions, combined with vision transformers that evaluate histopathology, provides unprecedented accuracy in survival risk stratification.

By proposing a holistic, multimodal framework, we highlight the necessity of bridging molecular genetics—such as BRCA mutations—with macroscopic tumor phenotypes and patient demographics. While challenges regarding data missingness, algorithmic bias, and computational costs remain prominent, the continuous evolution of cross-attention fusion mechanisms offers a robust path forward. Ultimately, synthesizing genetic vulnerability with environmental and visual biomarkers will dramatically reduce diagnostic errors, minimize false positives, and ensure that patients receive the highly personalized, life-saving treatments they require.

REFERENCES

Xie, Xinpeng, Li, Yuexiang, & Shen, Linlin (2018). *Active Learning for Breast Cancer Identification*. https://arxiv.org/pdf/1804.06670v1 https://arxiv.org/pdf/1804.06670v1

Farooq, Muhammad Shoaib, & Ilyas, Mehreen (2023). *Predicting environment effects on breast cancer by implementing machine learning*. https://arxiv.org/pdf/2309.14397v1 https://arxiv.org/pdf/2309.14397v1

Badré, Adrien, Zhang, Li, Muchero, Wellington, Reynolds, Justin C., & Pan, Chongle (2023). *Deep neural network improves the estimation of polygenic risk scores for breast cancer*. A. Badré, L. Zhang, W. Muchero, J.C. Reynolds, C. Pan (2021). Deep neural network improves the estimation of polygenic risk scores for breast cancer. Journal of Human Genetics, 66(4), 359-369. https://doi.org/10.1038/s10038-020-00832-7 https://doi.org/10.1038/s10038-020-00832-7

Tirunagari, Santosh, Poh, Norman, Abdulrahman, Hajara, Nemmour, Nawal, & Windridge, David (2015). *Breast Cancer Data Analytics With Missing Values: A study on Ethnic, Age and Income Groups*. https://arxiv.org/pdf/1503.03680v1 https://arxiv.org/pdf/1503.03680v1

Zuluaga-Gomez, Juan, Masry, Zeina Al, Benaggoune, Khaled, Meraghni, Safa, & Zerhouni, Noureddine (2019). *A CNN-based methodology for breast cancer diagnosis using thermal images*. https://arxiv.org/pdf/1910.13757v1 https://arxiv.org/pdf/1910.13757v1

Chakraborty, Debajyoti (2023). *Screening Mammography Breast Cancer Detection*. https://arxiv.org/pdf/2307.11274v1 https://arxiv.org/pdf/2307.11274v1

Good, Benjamin M., Loguercio, Salvatore, Griffith, Obi L., Nanis, Max, Wu, Chunlei, & Su, Andrew I. (2014). *The Cure: Making a game of gene selection for breast cancer survival prediction*. Good BM, Loguercio S, Griffith OL, Nanis M, Wu C, Su AI The Cure: Design and Evaluation of a Crowdsourcing Game for Gene Selection for Breast Cancer Survival Prediction JMIR Serious Games 2014;2(2):e7. https://doi.org/10.2196/games.3350 https://doi.org/10.2196/games.3350

Divyashree, BV, R, Amarnath, M, Naveen, & Kumar, G Hemantha (2018). *Novel approach to locate region of interest in mammograms for Breast cancer*. International Journal of Intelligent Systems and Applications in Engineering.(ISSN:2147-6799) Vol 6, No 3 (2018). https://doi.org/10.18201/ijisae.2018644775 https://doi.org/10.18201/ijisae.2018644775

Mondol, Raktim Kumar, Millar, Ewan K. A., Sowmya, Arcot, & Meijering, Erik (2024). *BioFusionNet: Deep Learning-Based Survival Risk Stratification in ER+ Breast Cancer Through Multifeature and Multimodal Data Fusion*. JBHI, 24 June 2024. https://doi.org/10.1109/JBHI.2024.3418341 https://doi.org/10.1109/JBHI.2024.3418341

Mondol, Raktim Kumar, Millar, Ewan K. A., Sowmya, Arcot, & Meijering, Erik (2024). *MM-SurvNet: Deep Learning-Based Survival Risk Stratification in Breast Cancer Through Multimodal Data Fusion*. https://arxiv.org/pdf/2402.11788v1 https://arxiv.org/pdf/2402.11788v1