Advancing Breast Cancer Research withArtificial Intelligence and Datasets

Basma Emhamed Dihoum [1] , Ahmed Ibrahim Suleiman[2]

[1]Aljafara University, College of Science and Natural Resources, Computer Department

[2]Faculty of Education Zawia , University of zawia

basmadihom@gmail.com

Abstract:

Globally, breast cancer, among the prevalent cancers in women, has experienced substantial advances in detection and treatment as a result of data-driven techniques. This review aims to encapsulate the recent progress and emerging trends in the study of breast cancer, particularly emphasizing non-image sequence data. This study delves into the latest developments in genomic, transcriptomic, and proteomic analyses that have revolutionized our understanding of breast cancer's molecular landscape. The review highlights how these data types are instrumental in identifying novel biomarkers, understanding tumor heterogeneity, and predicting therapeutic responses. Additionally, this study explores the integration of machine learning and artificial intelligence in analyzing sequence data, which has paved the way for personalized medicine and improved predictive models. By synthesizing findings from recent studies and clinical trials, this review presents acomprehensive overview of how non-image sequence data shapes the future of breast cancer research and therapy, offering insights into potential pathways for innovation and improved patient outcomes.

Keywords: Breast Cancer, Artificial Intelligence in Healthcare,Sequence Data.

1 INTRODUCTION

The American Cancer Society reports that breast cancer recently resulted in the deaths of 41,760 women and over 500 men. Four primary types of breast cancer as outlined by Benson, J. (2009). A benign tumor represents a slight alteration in breast tissue, typically considered non-harmful and non-cancerous. In-situ carcinoma, confined to the mammary duct lobule system, is less severe and treatable when caught early. The most severe, invasive carcinoma carries the potential to spread to other organs. Various detection approaches for breast cancer detection are available such as mammography, X-ray, ultrasound, Computed Tomography (CT), Portion Emission Tomography (PET), MRI, and breast temperature measurement, as noted by Hortobagyi, G. N. (1998). Pathological diagnosis, involving tissue removal and image analysis post-staining (commonly with Hematoxylin and Eosin), remains the gold standard for detection

Breast cancer diagnosis primarily employs either genomics or histopathological image analysis. microscopic, and Histopathological images views of breast tissue, play a crucial role in facilitating early cancer treatment.

Radio-genomics combines radiological and genetic data to enhance the diagnostic process. This approach allows for the evaluation of tissues at the molecular level,

contributing valuable insights for cancer prediction and detection. The essential difference between traditional imaging information and radio-genomics lies in the critical knowledge gap between tissue-level imaging and the assessment of underlying molecular and genetic disease indicators. While imaging, with its lower precision, may result in over- or under-treatment, radio-genomics proves more efficient than histopathological imaging. However, its limited adoption is attributed to its reliance on datasets demanding substantial processing capacity. As a result, only a few laboratories do radio-genomics research (Hortobagyi, 1998). This research paper delves into the realm of breast cancer detection and classification, emphasizing the role of DL models in comparison to conventional machine learning approaches. This study begins by scrutinizing the performance of various deep learning models, comparing them with classical counterparts, and discussing the metrics used to gauge their effectiveness. A significant part of this study is dedicated to understanding the crucial characteristics of breast cancer classification, and examining the methods used for their selection and extracted. This study also explores and enumerates the prominent datasets accessible for both gene sequencing and MRI imaging, providing insights into the methods employed for features selection and extraction. A distinctive aspect of this research is the comparative analysis between gene sequence data and image data, where This study dissects their respective strengths, limitations, and the unique challenges they pose in the context of breast cancer detection. This study provides a structured narrative that includes an extensive review of related work, a comprehensive methodology section, detailed findings, and a discussion that encapsulates the study's limitations and suggests avenues for future research. This study aim to offer a holistic view of the current landscape in breast cancer detection and classification, spotlighting the transformative potential of deep learning as a tool in oncological research and diagnostics.

2 BACKGROUND

2.1 What's breast cancer?

Breast cancer, a major global health issue, primarily affects women and occasionally

men and involves abnormal cell growth in breast tissues. Typically, this type of cancer originates in the lobules (milk-producing glands) or ducts (milk delivery pathways to the nipple), and less frequently in the stromal tissues, comprising the breast's fatty and connective components (Zagami, P 2022).

The disease's onset is often attributed to a mix of genetic and environmental influences. Key genetic risks involve mutations in genes like BRCA1 and BRCA2. Other factors increasing susceptibility include age, family breast cancer history, prolonged estrogen exposure, prior breast conditions, obesity, alcohol use, and radiation exposure. Breast cancer symptoms can vary but thickening in the breast or underarm area or often encompass lump, size or shape changes in the breast, skin alterations like dimpling, nipple changes such as retraction or discharge, and skin irritation on the breast. It's vital to recognize that not every change in breast tissue signifies cancer, yet any such changes warrant professional medical evaluation.

Diagnosing breast cancer typically involves physical exams, mammography, and biopsy procedures. The stage of cancer, determined by tumor size and spread, is crucial for formulating a treatment strategy. Treatment methods, depending on cancer type and stage, might include surgical interventions (like lumpectomy or mastectomy), radiation therapy, chemotherapy, hormone therapy, and precision therapy (So, J. Y 2022). With progress in treatment modalities, early detection has significantly improved survival rates.

Regular screening, such as mammograms, and awareness of breast health are crucial in the early detection of breast cancer. Ongoing research continues to offer new insights into this disease's prevention, diagnosis, and treatment, aiming for more personalized and effective therapies.

2.2 Breast cancer challenges

Breast cancer, as a complex and prevalent disease, presents a range of challenges that affect both patients and the healthcare system at large. The work of Hu, D (2022) and Titoriya, A. (2019) sheds light on these multifaceted issues. Some of the key challenges in managing and addressing breast cancer include:

- Early Detection: Despite advancements in screening, early detection is difficult. Mammograms, while primary tools, can miss cancers in dense breast tissue. There's also a need for heightened awareness about regular screenings and self-exams.

- Accurate Diagnosis: Differentiating benign from malignant lumps and pinpointing cancer type and stage is challenging, where errors can adversely affect treatment outcomes.
- Treatment Side Effects: Cancer treatments often have significant side effects. Balancing these effects with treatment efficacy is crucial.
- Personalized Medicine: Given breast cancer's heterogeneity, creating tailored treatments based on genetic factors and overall health is complex

but necessary.

- Metastasis: The spread of cancer to other body parts makes treatment more challenging and worsens prognosis.
- Drug Resistance: Breast cancer cells can become resistant to therapies over time, reducing treatment effectiveness for advanced cases.
- Psychological and Emotional Impact: The diagnosis and treatment can severely affect patients and their families, highlighting the need for supportive care.
- Financial Burden: Treatment costs can be overwhelming, causing financial stress and affecting access to quality care.
- Research and Development: Ongoing research is vital for understanding breast cancer mechanisms and developing better treatments.

- Global Disparities: There's a notable difference in breast cancer outcomes between developed and developing countries, largely due to varying access to healthcare resources.

Addressing these challenges requires a comprehensive approach, encompassing research, public health initiatives, better healthcare access, and patient support systems.

3 RELATED WORK

Numerous research projects have explored breast cancer detection using either imaging techniques or genomics. No study has combined both methodologies. Benson, J. (2009) studied various approaches for categorizing breast cancer through histological image analysis (HIA), with an emphasis on distinct Artificial Neural Network (ANN) architecture. Their review categorizes studies according to the datasets employed, which are listed chronologically. Their data show that ANNs became popular in HIA in about 2012, with ANNs and PNNs emerged as the predominant algorithms, although the majority of feature extraction experiments concentrated on textural and morphological aspects. Deep CNNs were discovered to be extremely effective in the early identification and diagnosis of breast cancer, resulting in better treatment outcomes. In addition, different algorithms have focused on predicting Noncommunicable Diseases (NCDs). Allugunti, V. R. (2022) compared the performance of different classification algorithms across eight NCD datasets.

A total of eight classification algorithms were employed alongside a 10-fold cross-validation technique, with their effectiveness assessed using the AUC as an accuracy metric. The researchers noted that the Non-Communicable Disease (NCD) datasets incorporated data with noise and irrelevant features. Algorithms like KNN, SVM, and NN were found to be resilient against the noise. They also

suggested that preprocessing techniques could address the issue of irrelevant attributes, thereby enhancing accuracy rates. In the domain of Natural Inspired Computing (NIC), algorithms inspired by insect behavior have been developed for diagnosing human diseases. Gautam R (2019) presented five insect-based NIC for diagnosing cancer and diabetes, observing high-performance levels in detecting various cancer types including lung, breast, ovarian and prostate. In the context of breast cancer, a blend of guided artificial colony (ABC) and neural networks was employed. The investigators also devised a successful method for identifying diabetes and leukemia, asserting that enhanced precision and promising outcomes arise when NICs are combined with alternative classification algorithms. Nevertheless, the study recognizes the necessity for additional research, particularly in discerning various stages of diabetes and cancer.

Alshayeji, M (2022), presents the effectiveness of Artificial Neural Networks (ANNs) in cancer classification, particularly in early stages, was demonstrated. Most NNs showed potential in identifying tumor cells, although the imaging technique required substantial digital resources for image pre-processing. In another review conducted by Hu, D (2022), a comprehensive examination of ML, DL, and data mining algorithms related to breast cancer prediction was undertaken, covering a diverse range of techniques in the field. While various algorithms like CNNs and Naïve Bayes have been employed in imaging techniques for breast cancer detection, In 2019, Titoriya highlighted gene mutation as a detection approach, giving particular attention to the gene prediction classification stage. This stage encompasses gene annotation, discovery, and mutation detection to discern the existence of cancer. The study identified a range of methods, such as regression, probability models, Support Vector Machines (SVMs), Neural Networks (NNs), and deep learning, for achieving this objective. They also highlighted the potential in exploring connections between nucleotide sequences and feature extraction, given the extensive data involved in DNA sequencing. Bahrami, M (2022) analyzed contemporary studies that apply DL to breast cancer using various imaging modalities. Classifying these investigations according to dataset, architecture, application, and evaluation methods, they concentrated specifically on deep learning frameworks applied to ultrasound, mammography, and MRI. The aim was to showcase the most recent developments in breast cancer imaging using computer-aided diagnosis systems based on deep learning (DLR-based CAD). This involved the utilization of private datasets and categorizations facilitated through convolutional neural networks (CNNs).

Building upon these surveys, this research will concurrently explore both genetic sequencing and imaging concurrently for breast cancer prediction. This dual approach aims to gather more comprehensive data that can assist in early diagnosis and treatment. Additionally, this study will offer insights and guidelines for researchers interested in pursuing this area of research.

Table1: Comparative results

Paper	Models/ Algorithm	Binary or Multiclass	Classes	Accuracy
(Xu, J, 2016)	CNN	Binary	Epithelial and Stromal	0.88
(Danaee, P 2017)	DNN	Multiclass	binary (Basal, Nonbasal)	0.83
(Romo-Bucheli 2017)	CNN, SVM, Random Forests,	Multiclass	Binary	0.97
(Rawat RR, 2018)	CNN	Multiclass	7 Cancer types	0.846
(Zemouri R, 2018)	Constructive	DNN	Low risk, intermediate risk or high risk	0.87
(Liu, Q 2019)	DNN + Attention mechanism	Binary	-	0.87
(Chen, H 2019)	GCN	Multiclass	-	0.919
(RAHEEM, A 2019)	BPNN	Binary	Mutant and non-mutant	0.998

(Elbashir, M, 2019)	CNN	Binary	Tumor or No	0.987
(Mostavi, M 2020)	CNN	Multiclass	-	0.956
(Jiang, P 2020)	DNN+SVM	Multiclass	Binary	0.94
(Alsaleem, M 2020)	CNN	Binary	Tumor or No	0.967
(Kumar A, 2020)	RF,SVM	Binary	Benign or Malignant	0.97
(Mostafa M,2020)	CNN	Binary	Normal or Abnormal	0.956
(Jebarani, P. E,2023)	GLCM+GWO+ DCLSTM	Binary	Normal or Abnormal	0.9925

4   Datasets available for gene sequencing

 Most private datasets used in breast cancer research originated from universities in the United States and the European Union, while only a handful of publicly available datasets were found that are free of charge.

Concerning gene data, although not as prevalent as imaging, a considerable amount of both private and public datasets are available for the study of both healthy and cancer-affected groups. The Cancer Genome Atlas stands out for its goal to catalog DNA changes across various cancers, including breast cancer, providing valuable genetic information across multiple cancer types to enhance understanding of cancer development. The METABRIC dataset is a crucial asset containing clinical information, gene expression data, Copy Number Variation profiles, and Single Nucleotide Polymorphism genotypes derived from breast cancer tumors in METABRIC study participants. The inclusion of the GEO database and the NCI Genomic Data Commons substantially expands the reservoir of gene-related data accessible for cancer research.

When compared to genetic data, a broader spectrum of datasets exists for

imaging data. The Wisconsin breast dataset, housed in the UCI repository and featuring data from digitized images of Fine Needle Aspirate samples of breast masses, stands out as one of the most frequently used imaging datasets.

Numerous other breast cancer imaging datasets are also available, predominantly public and free of cost. Large datasets, such as the DDSM, can produce reliable results on their own while combining smaller datasets can enhance the overall diversity of data.

5   Key Features Frequently Utilized in Breast Cancer Classification

Breast cancer features can be divided into two main categories: tumor characteristics and protein types/status. Tumor characteristics encompass essential factors like tumor size and grade, which vary depending on the specific type of cancer. The ER and PR status indicate the presence of specific proteins in the affected area. The count of positive lymph nodes provides a quantitative measure of cancer cells present and is reported accordingly. Metastasis refers to the dissemination of cancer to different organs, such as the lungs or bones. HER2 is another protein that plays a role in regulating the progression of breast cancer. The PAM50 score evaluates the risk of breast cancer spreading to other organs.

In terms of imaging data, a variety of general features are analyzed. These include the symmetry of the breast, its compactness, and the presence of concave points. Specific breast image characteristics are also examined. Marginal adhesion, for example, measures how cells at the periphery of the epithelial wall adhere to each other. Consistency in cell size is assessed, with a scale ranging from 0 (non-uniform) to 1 (uniform).

Clump thickness is evaluated to determine if cells are arranged in single or multiple layers, while assessing the consistency in cell size across the sample. Analyzing these imaging features provides a detailed view of breast characteristics, aiding the diagnostic process.

6   Top Methods for Selecting and Extracting Effective Features

Various feature extraction methods are explored in the literature. The Feature-based Strategy (FES), applied in some research, is geared towards identifying image displacements. It involves detecting and tracking specific features such as edges and corners across image frames. This process starts with the identification of features in consecutive images, simplifying the data, and then matching these features across frames to form motion vectors. Additionally, multi-level wavelet transformations are used by some researchers to extract time-frequency features from time series, dividing them into sub-series of different frequencies.

Convolutional Neural Networks (CNN) serve as another technique for feature selection and extraction. This method utilizes convolutional layers to extraction

features and fully connected layers for classification process. Typically, the process of feature extraction centers around the last hidden or convolutional layer after it has been flattened. For feature selection, techniques like XGboost and random forest are popular. These involve analyzing multiple layers of networks, with random forests assigning value to features based on how much they contribute to the purity of the network nodes. The most crucial features are those that substantially decrease node impurity. Conversely, Principal Component Analysis (PCA) is employed to condense the dimensions of data by generating new uncorrelated variables that systematicaly maximize variance, thus enhancing interpretability while minimizing information loss. When comparing genetic data with imaging data for breast cancer detection, each type presents its own set of strengths and challenges. Genetic data provides profound insights into the molecular aspects of cancer but can be complex to interpret and may have interpretational gaps. In contrast, imaging data, especially techniques like MRI, offers direct visual information but can vary in quality and be subject to interpretational differences. Combining both types of data can greatly enhance the accuracy and efficacy of the diagnostic process.

7   LIMITATIONS AND FUTURE RESEARCH

The integration of image and non-image sequence data in breast cancer research, while promising, faces several challenges. Firstly, the complexity and variability of data types require sophisticated analytical tools and algorithms, which are still in developmental stages. The quality and quantity of data available for analysis can also be a limiting factor, as inconsistent data collection methods across studiescan lead to difficulties in data harmonization. Additionally, there are privacy and ethical considerations related to the handling of sensitive patient data, which need to be rigorously addressed. The future of breast cancer diagnosis and treatment lies in the effective amalgamation of imaging and sequence data. Research should focus on developingadvanced computational models that can seamlessly integrate diverse data types, providing a more holistic view of the disease. Machine learning and AI techniqueshave the potential to be at the forefront of this endeavor in large datasets and uncover patterns that are not immediately apparent to human observers. Furthermore, the ethical use of patient data should be a priority, with research directed toward developing secure and transparent methods for data handling. The potential for early detection and personalized treatment plans is vast. With improved algorithms and data integration techniques, it may become possible to identify breast cancer at its earliest stages, even before it becomes detectable through conventional imaging.

   In conclusion, while challenges exist, the future research direction is clear. The integration of image and non-image sequence data in breast cancer research holds great promise for advancing our understanding of the disease, leading to earlier detection, and more personalized and effective treatment strategies.

8 DISCUSSION

The pivotal role of non-image sequence data in understanding breast cancer cannot be understated. While traditional imaging methods like mammography, MRI, and ultrasound have played crucial roles in detecting and monitoring breast cancer,they have limitations, particularly in deciphering the molecular intricacies of the disease. Non-image sequence data, encompassing genomic, transcriptomic, and proteomic analyses, offers a profound depth of understanding that transcends what can be visualized through imaging.

Genomic sequencing, for instance, allows for the identification of specific mutations and genetic alterations in breast cancer cells, which might not be discernible through imaging. These genetic markers can be crucial for understanding the aggressiveness of the tumor, its potential to metastasize, and its responsiveness to specific treatments. Transcriptomic and proteomic analyses further complement this by providing insights into the active biological processes within the tumor cells. This level of molecular profiling is vital for uncovering mechanisms of drug resistance, tumor progression, and recurrence, which are often invisible toconventional imaging techniques.

Furthermore, the integration of machine learning and artificial intelligence with sequence data analysis is revolutionizing breast cancer prognosis and treatment

planning. These technologies enable the analysis of vast datasets to identify patterns and correlations that would be impossible for the human eye to discern. This approach is particularly valuable in cases where imaging results are ambiguous or in early-stage cancers where imaging might be less effective.

In summary, non-image sequence data provides an invaluable tool for understanding-ing the complexities of breast cancer at a molecular level. It offers a complementary perspective to imaging, allowing for a more comprehensive and precise approach to diagnosis, prognosis, and treatment. As such, the medical community needs to continue advancing and integrating these data-driven techniques to enhance our understanding and management of breast cancer.

9 CONCLUSION

The exploration of breast cancer through the lens of both imaging and non-image sequence data has opened new frontiers in understanding and managing this complex disease. While imaging provides a visual assessment of the physical characteristics of tumors, sequence data delves into the molecular and genetic underpinnings, offering a complementary and more nuanced perspective. This integrative method holds the potential to transform breast cancer diagnosis, prognosis, and treatment, moving towards more personalized and effective healthcare strategies.

Despite the challenges in data integration and analysis, advancements in machine learning and artificial intelligence offer promising solutions. These technologies enable the synthesis of diverse data types, providing insights that are greater than the sum of their parts. The continued development and refinement of these tools will be crucial for realizing the full potential of this integrative approach. Moreover, the ethical and secure handling of patient data remains a priority. Ensuring patient privacy and data integrity is not only a legal imperative but also a cornerstone in maintaining public trust and advancing medical research.

In summary, the combination of imaging and non-image sequence data represents a significant step forward in the battle against breast cancer. It promises earlier detection, improved prognostic accuracy, and more tailored treatment options. As research continues to evolve in this field, it holds the promise of transforming the landscape of breast cancer care, ultimately leading to better outcomes for patients worldwide.

REFERENCES

Benson, J. R., Jatoi, I., Keisch, M., Esteva, F. J., Makris, A., & Jordan, V. C. (2009). Early breast cancer. The Lancet, 373(9673), 1463-1479.

Hortobagyi, G. N. (1998). Treatment of breast cancer. New England Journal of Medicine, 339(14), 974-984.

Zagami, P., & Carey, L. A. (2022). Triple-negative breast cancer: Pitfalls and progress. NPJ Breast Cancer, 8(1), 95.

So, J. Y., Ohm, J., Lipkowitz, S., & Yang, L. (2022). Triple-negative breast cancer (TNBC): Non-genetic tumor heterogeneity and immune microenvironment: Emerging treatment options. Pharmacology & therapeutics, 108253.

Hu, D., Li, Z., Zheng, B., Lin, X., Pan, Y., Gong, P., ... & Wang, L. (2022). Cancer-associated fibroblasts in breast cancer: Challenges and opportunities. Cancer Communications, 42(5), 401-434.

Titoriya, A., & Sachdeva, S. (2019, November). Breast cancer histopathology image classification using AlexNet. In 2019 4th International Conference on Information Systems and computer networks (ISCON) (pp. 708-712). IEEE.

Bahrami, M., Sadeghi, A., Mosavizadeh, R., & Masoumy, M. (2022). Challenges

of meeting the palliative care needs of colorectal cancer patients in Iran: A qualitative research. Iranian Journal of Nursing and Midwifery Research, 27(4), 325-330.

(Allugunti, V. R 2022) Allugunti, V. R. (2022). Breast cancer detection based on thermographic images using machine learning and deep learning algorithms. International Journal of Engineering in Computer Science, 4(1), 49-56.

Gautam R, Kaur P, Sharma M. A comprehensive review on nature inspired computing algorithms for the diagnosis of chronic disorders in human beings. Prog Artif Intell 2019;8:401–24.

Alshayeji, M. H., Ellethy, H., & Gupta, R. (2022). Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach. Biomedical Signal Processing and Control, 71, 103141.

Punitha, S., Amuthan, A., & Joseph, K. S. (2019). Enhanced Monarchy Butterfly Optimization Technique for effective breast cancer diagnosis. Journal of Medical Systems, 43, 1-14.

Umamaheswari, T. S., & Sumathi, P. (2022). Enhanced firefly algorithm (EFA) based gene selection and adaptive neuro neutrosophic inference system (ANNIS) prediction model for detection of circulating tumor cells (CTCs) in breast cancer analysis (Retraction of 10.1007/S10586-022-03845-3, 2022)(Retracted article. See DEC, 2022).

Shen, J., Shi, J., Luo, J., Zhai, H., Liu, X., Wu, Z., ... & Luo, H. (2022). Deep learning approach for cancer subtype classification using high-dimensional gene expression data. BMC bioinformatics, 23(1), 1-17.

RAHEEM, A. K. A., & DHANNOON, B. N. (2019). Predication and classification of cancer using sequence alignment and back propagation algorithms in brca1 and brca2 genes. International Journal of Pharmaceutical Research, 11(1).

Mostavi, M., Chiu, Y. C., Huang, Y., & Chen, Y. (2020). Convolutional neural network models for cancer type prediction based on gene expression. BMC medical genomics, 13, 1-13.

Liu, Q., & Hu, P. (2019). Association analysis of deep genomic features extracted by denoising autoencoders in breast cancer. Cancers, 11(4), 494.

Chen, H., Gao, M., Zhang, Y., Liang, W., & Zou, X. (2019). Attention-based multi-NMF deep neural network with multimodality data for breast cancer prognosis model. BioMed research international, 2019.

Jiang, P., Huang, S., Fu, Z., Sun, Z., Lakowski, T. M., & Hu, P. (2020). Deep graph embedding for prioritizing synergistic anticancer drug combinations. Computational and structural biotechnology journal, 18, 427-438.

Danaee, P., Ghaeini, R., & Hendrix, D. A. (2017). A deep learning approach for cancer detection and relevant gene identification. In Pacific symposium on biocomputing 2017 (pp. 219-229).

Romo-Bucheli, D., Janowczyk, A., Gilmore, H., Romero, E., & Madabhushi, A. (2017). A deep learning based strategy for identifying and associating mitotic activity with gene expression derived risk categories in estrogen receptor positive breast cancers. Cytometry Part A, 91(6), 566-573.

El-Bendary, N., & Belal, N. A. (2020). A feature-fusion framework of clinical, genomics, and histopathological data for METABRIC breast cancer subtype classification. Applied Soft Computing, 91, 106238.

Alsaleem, M. A., Ball, G., Toss, M. S., Raafat, S., Aleskandarany, M., Joseph, C., ... & Rakha, E. (2020). A novel prognostic two-gene signature for triple negative breast cancer. Modern Pathology, 33(11), 2208-2220.

Lancashire, L. J., Powe, D. G., Reis-Filho, J. S., Rakha, E., Lemetre, C., Weigelt, B., ... & Ball, G. R. (2010). A validated gene expression profile for detecting clinical outcome in breast cancer using artificial neural networks. Breast cancer research and treatment, 120, 83-93.

Rawat, R. R., Ruderman, D., Macklin, P., Rimm, D. L., & Agus, D. B. (2018). Correlating nuclear morphometric patterns with estrogen receptor status in breast cancer pathologic specimens. NPJ Breast Cancer, 4(1), 32.

Zhang, D., Zou, L., Zhou, X., & He, F. (2018). Integrating feature selection and feature extraction methods with deep learning to predict clinical outcome of breast cancer. Ieee Access, 6, 28936-28944.

Beykikhoshk, A., Quinn, T. P., Lee, S. C., Tran, T., & Venkatesh, S. (2020). DeepTRIAGE: interpretable and individualised biomarker scores using attention mechanism for the classification of breast cancer sub-types. BMC medical genomics, 13, 1-10.

Cruz-Roa, A., Gilmore, H., Basavanhally, A., Feldman, M., Ganesan, S., Shih, N., ... & González, F. (2018). High-throughput adaptive sampling for whole-slide histopathology image analysis (HASHI) via convolutional neural networks: Application to invasive breast cancer detection. PloS one, 13(5), e0196828.

Paul, A., & Mukherjee, D. P. (2015). Mitosis detection for invasive breast cancer grading in histopathological images. IEEE transactions on image processing, 24(11), 4041-4054.

Rawat, R. R., Ruderman, D., Macklin, P., Rimm, D. L., & Agus, D. B. (2018). Correlating nuclear morphometric patterns with estrogen receptor status in breast cancer pathologic specimens. NPJ Breast Cancer, 4(1), 32.

Xu, J., Luo, X., Wang, G., Gilmore, H., & Madabhushi, A. (2016). A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images. Neurocomputing, 191, 214-223.

Kumar, A., Singh, S. K., Saxena, S., Lakshmanan, K., Sangaiah, A. K., Chauhan, H., ... & Singh, R. K. (2020). Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer. Information Sciences, 508, 405-421.

Mostavi, M., Chiu, Y. C., Huang, Y., & Chen, Y. (2020). Convolutional neural network models for cancer type prediction based on gene expression. BMC medical genomics, 13, 1-13.

Zemouri, R., Omri, N., Morello, B., Devalland, C., Arnould, L., Zerhouni, N., & Fnaiech, F. (2018). Constructive deep neural network for breast cancer diagnosis. IFAC-PapersOnLine, 51(27)