Advancing Histopathology-Based Breast Cancer Diagnosis: Insights into Multi-Modality and Explainability

Faseela Abdullakutty[1*], Younes Akbari[1*], Somaya Al-Maadeed[1*], Ahmed Bouridane[2] and Rifat Hamoudi[2]

[1*] Qatar University, Qatar.
[2] University of Sharjah, United Arab Emirates.

*Corresponding author(s). E-mail(s): faseela.abdullakutty@qu.edu.qa;
akbari_younes@yahoo.com; s_alali@qu.edu.qa;
Contributing authors: abouridane@sharjah.ac.ae;
rhamoudi@sharjah.ac.ae;

Abstract

It is imperative that breast cancer is detected precisely and timely to improve patient outcomes. Diagnostic methodologies have traditionally relied on unimodal approaches; however, medical data analytics is integrating diverse data sources beyond conventional imaging. Using multi-modal techniques, integrating both image and non-image data, marks a transformative advancement in breast cancer diagnosis. The purpose of this review is to explore the burgeoning field of multimodal techniques, particularly the fusion of histopathology images with non-image data. Further, Explainable AI (XAI) will be used to elucidate the decision-making processes of complex algorithms, emphasizing the necessity of explainability in diagnostic processes. This review utilizes multi-modal data and emphasizes explainability to enhance diagnostic accuracy, clinician confidence, and patient engagement, ultimately fostering more personalized treatment strategies for breast cancer, while also identifying research gaps in multi-modality and explainability, guiding future studies, and contributing to the strategic direction of the field.

Keywords: Breast Cancer detection, multi-modality, explainability

1

1 Introduction

In the realm of breast cancer diagnosis, the convergence of multi-modal techniques, amalgamating both image and non-image data [1], heralds a transformative approach with profound implications for disease detection and characterization. As a leading cause of mortality among women globally, the precise and timely diagnosis of breast cancer remains imperative for optimizing patient outcomes. While traditional diagnostic methodologies [2] have historically relied heavily on uni-modal approaches, the evolving landscape of medical data analytics underscores the significance of integrating diverse data sources beyond conventional imaging modalities [3].

Figure 1 illustrates a generic model for breast cancer diagnosis within the Computer-Aided Detection (CAD) framework. As depicted in Figure 2, breast cancer detection can be performed using various data types, employing either unimodal or multimodal approaches. The process initiates with data pre-processing, followed by feature extraction. To enhance the learning of feature representations from image data, segmentation may be conducted prior to feature extraction. Subsequently, the detection model is applied to generate a diagnosis from the processed data. Based on this diagnosis, further analyses are performed, including sub-type classification, grade classification, recurrence and metastasis prediction, as well as the incorporation of crowdsourcing and human-in-the-loop methodologies. These steps culminate in a final decision that informs subsequent treatment and monitoring strategies.
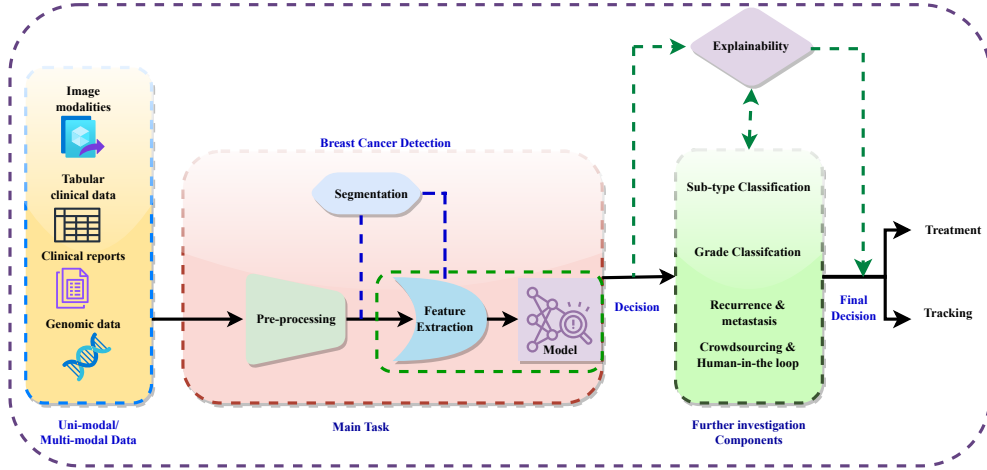

Fig. 1: A generic representation of breast cancer diagnosis

This review endeavours to illuminate the burgeoning field of multi-modal techniques in breast cancer diagnosis context, placing particular emphasis on the fusion of heterogeneous data streams encompassing both image and non-image modalities. Beyond the confines of traditional imaging modalities, such as mammography, magnetic resonance imaging (MRI), ultrasound, and positron emission tomography (PET),

multi-modal approaches leverage a plethora of non-image data types including genetic markers, proteomic profiles, clinical parameters, and patient demographics [4, 5]. By harnessing the complementary insights gleaned from these diverse data modalities, multi-modal techniques offer a multifaceted understanding of breast cancer biology and pathology, transcending the limitations of uni-modal approaches.
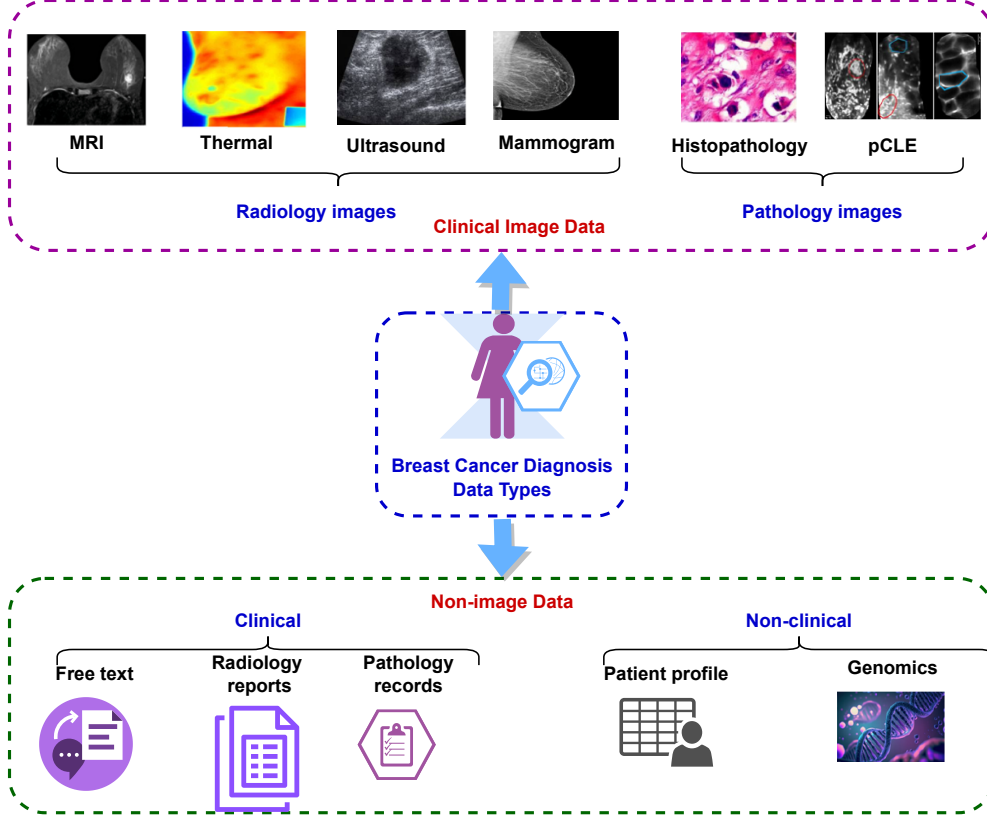

Fig. 2: Types of breast cancer diagnosis data

Furthermore, alongside the integration of multi-modal data, the exigency for explainability in breast cancer diagnosis emerges as a pivotal consideration. As machine learning and artificial intelligence algorithms increasingly permeate diagnostic workflows, the interpretability and transparency of decision-making processes assume paramount importance. Explainable AI (XAI) techniques endeavor to demystify the opaque nature of complex algorithms, elucidating the rationale behind diagnostic decisions and enhancing the interpretability of diagnostic outcomes [5]. In the context of breast cancer diagnosis, explainability not only engenders clinician confidence

in decision support systems but also fosters patient understanding and engagement, empowering informed decision-making and facilitating personalized treatment strategies.

Table 1: Latest reviews on breast cancer diagnosis in various contexts.

Author	Year	Main Discussion	Datasets	Multi modality	XAI
Abo-El-Rejalet al.[4]	2024	Segmentation	×	×	×
Brodhead et al.[6]	2024	Imaging characteristics	×	×	×
Hussain et al.[5]	2024	Breast cancer risk prediction	×	✓	✓
Luo et al.[7]	2024	Breast Cancer Imaging	×	✓	×
Rautela et al.[8]	2024	Computational techniques for breast cancer	×	✓	×
Singh et al.[9]	2024	Breast Cancer Screening and Detection using Artificial Intelligence and Radiomics	×	✓	×
Thakur et al .[10]	2024	Identification and of breast cancer through medical image modalities	✓	✓	×

Table 1 presents recent reviews on breast cancer diagnosis across various contexts. However, these reviews often overlook multi-modality and explainability, treating them as future research directions rather than discussing existing methods. Additionally, there is a lack of focus on histopathology and frameworks that combine histopathology with non-image data for breast cancer detection. In light of these observations, this review addresses multi-modal datasets, including histopathology and other non-image data, explores multi-modal techniques utilizing these datasets, and examines explainable multi-modal methods in histopathology-based breast cancer diagnosis.

The major contributions of this article are:

- A detailed investigation of multi-modal datasets, including those that incorporate histopathology and non-image data, which are frequently overlooked in existing literature.
- A discussion on multi-modal techniques that utilize the aforementioned datasets, offering insights into their application and effectiveness in breast cancer diagnosis.
- An investigation of explainable multi-modal methods specifically within the context of histopathology-based breast cancer diagnosis, addressing a critical gap in current research.
- Identification research gaps in multi-modality and explainability, identifying key areas for future study and contributing to the strategic direction of the field.

2 Breast Cancer Diagnosis: An Overview

Breast cancer is among the most prevalent malignancies impacting women globally, presenting substantial challenges for public health and individual well-being. The importance of early diagnosis cannot be overstated, as it is pivotal in enhancing treatment efficacy and elevating survival rates. This overview delves into the diverse tasks

involved in breast cancer diagnosis. By critically examining the current methodologies and advancements in breast cancer detection [11], strategies for ensuring prompt, precise, and effective diagnostic processes can be elucidated, thereby fostering improved patient care and clinical outcomes.

2.1 Tasks

Breast cancer diagnosis [12] involves several tasks that can utilize image and non-image data as shown in Figure. 3. Breast cancer detection entails pinpointing signs of cancer within breast imaging data, such as mammograms or ultrasound scans. By harnessing machine learning algorithms, it's possible to analyze these images to identify potentially suspicious areas or anomalies that could suggest the presence of tumours. These advanced techniques [13] offer a more efficient and potentially more accurate method for detecting early signs of breast cancer, providing valuable insights for healthcare professionals in their diagnostic process.

Malignancy classification [14] is the process of determining whether detected abnormalities are malignant, indicating cancer, or benign, meaning they are non-cancerous. This step is vital for guiding the subsequent treatment plan. Machine learning models can assist in this classification by analyzing features derived from imaging data, including characteristics like shape, texture, and intensity. By training these models on large datasets, they can provide predictions on the probability that an abnormality is cancerous, aiding healthcare professionals in making informed decisions regarding patient care [15].
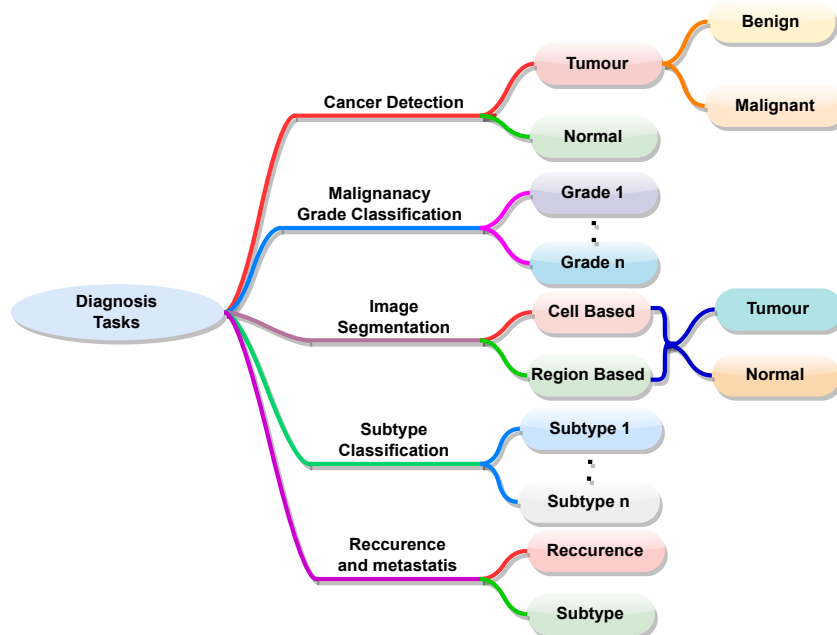

Fig. 3: Breast Cancer diagnosis tasks

Subtype classification is a crucial process in understanding breast cancer, as it encompasses a spectrum of diseases, each with unique traits and outcomes [16]. This step involves dividing breast cancer cases into specific subtypes like hormone receptor-positive, HER2-positive, or triple-negative breast cancer, which are known to have varying responses to treatments and differing prognoses. By categorizing cases into these subtypes, medical professionals can tailor treatment plans more effectively [17]. Machine learning models play a role in this by analyzing genomic data, gene expression profiles, and clinical information to predict the subtype, facilitating personalized and targeted therapeutic approaches.

Table 2: Recent research in breast cancer diagnosis including different tasks

Method	Dataset	Modality	Task
Classifier-combined method [14]	Proprietary	MRI	Grade Classification
DeepBreastCancerNet [18]	BUSI [19], Ultrasound Image dataset[20]	Ultrasound	Detection
DSCCN [16]	TCGA [21]	multi-omics	Sub-type Classification
EMDCOC [22]	BreakHis [23] IR Thermal Images [24]	Histopathology, IR thermal images	Detection
Ensemble CNN [15]	Databiox [25]	Histopathology	Grade Classification
histogram K-means segmentation [26]	BreakHis [23]	Histopathology	Segmentation
Hybrid CNN [27]	Mini-DDSM [28], BUSI [19]	Mammogram, Ultrasound images	Detection
Hybrid CNN-LSTM [29]	BreakHis [23]	Histopathology	Grade Classification
KAMnet [30]	Proprietary	Ultrasound	Detection
moBRCA-net [17]	TCGA [21]	Multi-omics,	Sub-type Classification
Multi-modal fusion [31]	TCGA [21]	WSI, Gene Expression	Detection
optimized LSTM with U-net segmentation[32]	MIAS [33]	Mammogram	Segmentation
Prediction model for distant metastasis[34]	Proprietary	Clinical Data	Reccurence and metastatis
recurrence prediction [35]	WPBC	Clinical Data	Recurrence and metastasis
Semantic Segmentation [36]	CBIS-DDSM [37], MIAS [33]	Mammogram	Segmentation
Unet3+ [38]	Proprietary	Ultrasound	Segmentation
Yolo-Based Model [39]	CBIS-DDSM [37], Inbreast [40], Proprietary	Mammogram	Detection

Image segmentation [13] involves dividing an image into cell segmentation and distinct segments or regions of interest. Within the realm of breast cancer diagnosis, segmentation helps to demarcate the boundaries of tumours or suspicious lesions in breast imaging data [41]. This process is critical for precisely measuring tumour size

and shape, and it lays the groundwork for further analyses, including tumour volume estimation or extracting quantitative features. Machine learning algorithms, especially deep learning models like convolutional neural networks (CNNs), have demonstrated strong capabilities in automatically segmenting breast lesions from medical images [42], offering a powerful tool to enhance the accuracy and efficiency of breast cancer diagnosis.

Predicting cancer recurrence and metastasis [43] is a crucial aspect of breast cancer management, extending beyond initial diagnosis and treatment. This task involves assessing the risk of the cancer returning or spreading to other parts of the body. Machine learning models can combine multiple types of data—such as imaging, genomic information, clinical variables (like patient demographics and medical history), and treatment records—to estimate the likelihood of recurrence or metastasis [35]. These predictions are valuable for clinicians, allowing them to customize follow-up care and create personalized treatment plans for breast cancer patients, ultimately enhancing patient outcomes and reducing the risk of adverse events. It should be noted that the tasks should be combined and integrated to have an accurate system. For example, cancer detection for subtype classification should use the tasks of cancer segmentation and grading tasks and this process can improve the task of subtype classification [44].

Table. 2 presents a summary of recent research advancements in breast cancer diagnosis across various tasks. A significant observation is the predominance of unimodal approaches in current methodologies. While some existing multimodal methods incorporate different types of imaging, such as ultrasound and mammography, the integration of image data with non-image data remains significantly underexplored. In particular, the fusion of histopathology images with non-image data, including textual and clinical information, represents a largely untapped area. The potential benefits of this integration are substantial. By combining histopathological imaging with comprehensive clinical and textual data, and leveraging advanced machine learning techniques, there is a strong potential to enhance the accuracy and efficiency of breast cancer diagnosis, prognosis, and treatment planning. This holistic approach could lead to significant advancements in personalized medicine and improved patient outcomes.

2.2 Datasets

The dataset used for breast cancer diagnosis encompasses both clinical image data and non-image data [45], as illustrated in Figure 2. The clinical image data comprise radiology and pathology images. Radiology images encompass modalities such as MRI, CT, thermal imaging, mammograms, and ultrasound, while pathology images include histopathology and pCLE [3]. The non-image data can be subdivided into clinical and non-clinical categories. Clinical data encompass radiology reports, pathology reports, including laboratory results, and narrative descriptions of patient status. Non-clinical data comprise patient profiles containing demographic information, patient history, age, other non-clinical details, and genomic data [46].

Additionally, non-image data are further classified into structured and unstructured categories. Radiology reports and narrative descriptions of patient status fall under unstructured data, while recorded pathology reports and patient profiles are

considered structured data [47]. Despite the abundance of both image and non-image datasets related to breast cancer detection, this paper focuses specifically on histopathology-based datasets, examining them in a multi-modal context. Table. 3, lists the existing public datasets in breast cancer detection, based on histopathology. It is evident from the table that the number of multi-modal datasets is much less compared to the unimodal datasets. Also, the sample size is low in most of these datasets.

The landscape of breast cancer histopathology research is enriched by a diverse array of datasets, each offering unique features and clinical insights. Uni-modal datasets, such as BRACS [48] and BreCaHAD [49], focus on a single type of data. The BRACS dataset provides 547 Whole-Slide Images (WSIs) and 4539 Regions Of Interest (ROIs), meticulously annotated by three board-certified pathologists. This dataset categorizes lesions into types such as Normal, Pathological Benign, Usual Ductal Hyperplasia, Flat Epithelial Atypia, Atypical Ductal Hyperplasia, Ductal Carcinoma in Situ, and Invasive Carcinoma. Similarly, the BreCaHAD dataset includes 162 histopathology images focusing on malignant cases, classified into mitosis, apoptosis, tumour nuclei, non-tumour nuclei, tubule, and non-tubule, thus facilitating comprehensive analyses and validation of diagnostic methods.

In contrast, multi-modal datasets integrate various data types to provide a more comprehensive view of breast cancer pathology. The TCGA-BRCA [21] dataset, for instance, combines gene expression data, copy number variations (CNVs), and pathological images from 1098 breast cancer patients. This multi-dimensional approach allows for a deeper understanding of the molecular and histological characteristics of breast cancer. Similarly, the IMPRESS dataset includes Hematoxylin and Eosin (H&E) and immunohistochemistry (IHC) stained WSIs from 126 patients, along with clinical data and biomarker annotations. The Post-NAT-BRCA38 dataset [50] offers 96 WSIs along with detailed clinical information, including estrogen receptor (ER) status, progesterone receptor (PR) status, and human epidermal growth factor receptor 2 (HER2) status. These multi-modal datasets enable researchers to explore the interplay between genetic, molecular, and histological data, driving advancements in personalized breast cancer diagnosis and treatment.

Table 3: Multi-Modal Datasets public for Breast Cancer Diagnosis Featuring Histopathology Images

Dataset	Year	Size	Modalities
Post-NAT-BRCA [50]	2019	96	WSI ,Clinical data
CPTAC-BRCA [51]	2020	642	WSI, clinical, proteomic, genomic data
Pathological EMR [52]	2021		WSI, patient profile
BCNB [53]	2022	1,058	Clinical Data
IMPRESS [54]	2023	126	WSI ,Clinical data
GTEx-Breast dataset [55]	2023	894	WSI, pathology notes
TCGA-BRCA dataset [21]	2023	1098	WSI, gene expression, CNV

The IMPRESS dataset [54] consists of 126 breast H&E WSIs from 62 female patients with HER2-positive breast cancer and 64 female patients with triple-negative breast cancer, all of whom underwent neoadjuvant chemotherapy followed by surgical excision. It includes immunohistochemistry (IHC) stained WSIs of the same slides, along with corresponding scores. All slides were scanned using a Hamamatsu scanner at $20\times$ magnification. The dataset also provides clinical data for both patient groups, including age, tumour size, and annotations for biomarkers such as PD-L1, CD-8, and CD-163. The GTEx-Breast dataset [55] is part of the Genotype-Tissue Expression (GTEx) project, which offers gene expression data across 44 human tissues. It includes 894 breast tissue histology images, comprising 306 WSIs of female breast tissue and 588 WSIs of male breast tissue, collected from the central subareolar region of the right breast at various centers in the United States. The images are accompanied by brief pathology notes and an annotation file with detailed sample information.

The CPTAC-BRCA dataset [51], from the Clinical Proteomic Tumor Analysis Consortium, includes 642 WSIs from 134 patients with breast invasive carcinoma, scanned at $20\times$ magnification. The images are available in two resolutions: 0.25 mum/pixel and 0.5 mum/pixel. The dataset is accompanied by comprehensive clinical, proteomic, and genomic data. The BCNB dataset [53], or Early Breast Cancer Core-Needle Biopsy WSI Dataset, is the only publicly available collection of breast histopathological WSIs from Asia. It contains 1,058 WSIs from 1,058 breast cancer patients in China, scanned with an Iscan Coreo pathological scanner. Tumor regions in each image are annotated by two pathologists. The dataset also includes extensive clinical data such as patient age, tumour size, histological and molecular subtypes, number of lymph node metastases, and HER2, ER, and PR status.

3 Exploring multi-modality

Multi-modal techniques are increasingly significant in histopathology-based breast cancer detection due to their capability to enhance diagnostic accuracy, provide comprehensive insights, and improve patient outcomes. These techniques integrate various types of data and analytical methods, offering a more robust framework for detecting and characterizing breast cancer. Multimodal data fusion combines information from different modalities to improve decision-making processes, which is particularly useful in medical fields where single observations may yield diverse interpretations. For instance, while isocitrate dehydrogenase (IDH1) mutation status and histological profiles individually contribute to understanding patient outcomes, their combined analysis has been crucial in revising the WHO classification of diffuse gliomas. Similarly, in breast cancer detection, artificial intelligence (AI) provides an automated and objective means to incorporate complementary information and clinical context from diverse datasets, thereby enhancing predictive accuracy.

As shown in Figure. 4, the multi-modal fusion can be categorised as stage-based and method-based techniques. Stage-based fusion strategies can be further categorized into early, late, and intermediate fusion approaches [56], each offering unique advantages in breast cancer detection. This approach is particularly beneficial when unimodal data

are noisy or incomplete, as integrating redundant information from other modalities can improve the robustness and precision of predictions.
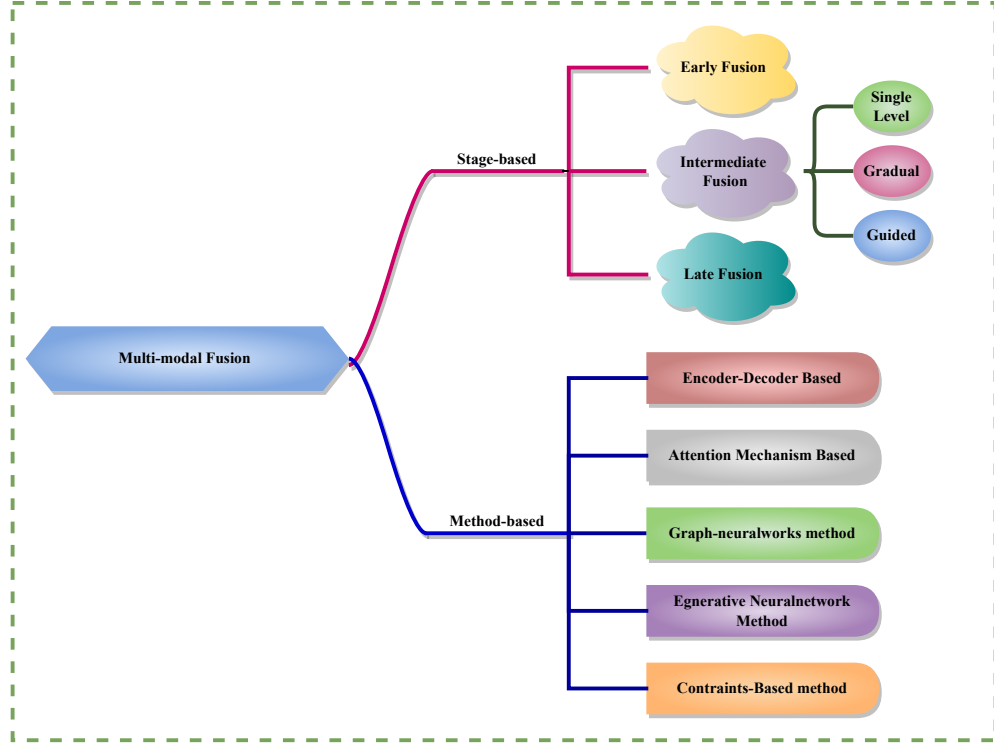

Fig. 4: Multi-modal approaches

Multi-modal fusion approaches [57] include encoder-decoder methods, which combine feature extraction, fusion, and decision-making processes into a single model, making them efficient in tasks like video captioning and object detection. Attention mechanism methods use mechanisms like co-attention and cross-attention to enhance each modality with information from other modalities, allowing the model to fuse features and learn interdependencies among them. Graph Neural Network methods use GNN to capture long-range dependencies among different modalities, categorizing tasks into different classes based on data types. Generative Neural Network methods include models like VAE-based adversarial frameworks, which reduce distance differences between unimodal representations and are crucial for tasks like text-conditional image generation and image style transfer. Constrained-based methods involve innovative approaches like channel-exchanging-networks, which dynamically exchange channels in different modal sub-networks based on individual channel importance, but are limited to homogeneous data.

Multi-modal techniques can improve diagnostic accuracy by combining different data modalities such as histopathological images, molecular profiles, and clinical data.

By integrating these data, they can differentiate between cancer subtypes, assess tumour heterogeneity, and predict potential treatment responses. They also contribute to better prognostication and personalized medicine by providing a holistic view of tumour biology and patient condition. Advanced imaging and computational tools, such as machine learning and artificial intelligence (AI), have revolutionized the analysis of histopathological data, automating the detection and classification of cancerous cells, extracting and analyzing complex patterns, and providing decision support to pathologists. In research, multi-modal techniques facilitate a deeper understanding of breast cancer mechanisms, leading to the discovery of new therapeutic targets and biomarkers. In clinical settings, these techniques enhance diagnostic workflows and support real-time decision-making, ultimately improving patient care.

Multi-modal techniques offer a more accurate, comprehensive, and personalized approach to breast cancer diagnosis and treatment, but they face challenges such as data standardization, computational resources management, and interdisciplinary collaboration. Future advancements in technology and computational methods are expected to address these challenges, making multimodal techniques more effective and widely adopted in clinical practice. However, challenges persist, such as the richness of feature representation [52] in images and the inadequacy of information fusion, which can lead to the loss of high-dimensional information and partially missing data in real-world scenarios. Each modality within multimodal data possesses distinct characteristics, adding to the complexity of heterogeneous data and further complicating multimodal fusion methods.

3.1 Multi-Modal Breast Cancer Diagnosis: Combining Histopathology and Non-Image Modalities

The integration of multi-modal approaches in breast cancer diagnosis, including histopathology and non-image modalities, improves diagnostic accuracy, provides a comprehensive understanding of the disease, improves personalized treatment planning, facilitates early detection and timely intervention, potentially improving patient outcomes, and promotes interdisciplinary collaboration among specialists. This approach reduces the likelihood of misdiagnosis, provides a more comprehensive understanding of tumor biology and patient health, and facilitates early detection and timely intervention, ultimately advancing clinical research. Table. 4 shows recent multi-modal research in breast cancer diagnosis.

Combining heterogeneous data without losing information from high-dimensional images has posed a significant challenge in data fusion. To address this issue, the authors of [52] proposed a multi-modal fusion technique that increased the dimensionality of structured data to align with the high-dimensional image features from histopathology WSI. They employed VGG-16 as the backbone model for image feature extraction and utilized a denoising autoencoder to enhance the dimensionality of the structured clinical data. The extracted features were then flattened and combined through concatenation before being input into fully connected layers for classification. This model successfully classified cases as benign or malignant, with its performance evaluated using the pathological electronic medical record (EMR) dataset. PathLDM [58], is a text-conditioned Latent Diffusion Model designed to improve histopathology

image generation by leveraging contextual information from pathology text reports. The model utilized Generalized Pathological Text (GPT) to distill and summarize complex text reports, thereby establishing an effective conditioning mechanism. The authors achieved a state-of-the-art Fréchet Inception Distance (FID) score of 7.64 for text-to-image generation on the TCGA-BRCA dataset, surpassing the closest text-conditioned competitor. The study compared PathLDM against alternative methods such as Moghadam et al., Medfusion, and Stable Diffusion, demonstrating its superiority in generating high-quality histopathology images conditioned on text.

Authors of [59] presented a method for improving mitosis detection in histopathology images using large vision-language models, combining image captioning and visual question-answering tasks with pre-trained models that integrate visual features and natural language. The inclusion of metadata, such as tumor and scanner types, into the question prompts significantly enhanced prediction accuracy. This method outperformed baseline models, including single-modality models and the vision-language model CLIP, demonstrating superior mitosis detection. Lu et al. (2023) [60] introduced the MI-Zero framework, which utilized contrastively aligned image and text models for zero-shot transfer on gigapixel histopathology whole slide images. Reformulating zero-shot transfer through multiple instance learning for large images, the text encoder was pre-trained with over 550,000 pathology reports and in-domain text corpora. Evaluated on three WSI datasets from Brigham and Women's Hospital, MI-Zero used independent datasets to prevent information leakage and employed a graph-based representation considering the spatial positions of each patch for slide-level prediction scores, significantly advancing cancer subtype classification accuracy and robustness.

A bi-phase model [61] was developed for predicting breast cancer prognosis, integrating multi-modal data from genomic information, histopathology images, and clinical details. The model was evaluated using two datasets: METABRIC (1980 patients) and TCGA-BRCA (1080 patients). The methodology employed a fusion strategy that began with feature extraction from each uni-modal dataset using separate SiGaAtCNNs. The extracted hidden features were then concatenated with input features to form stacked features. These stacked features were subsequently fed into random forest classifiers enhanced with adaptive boosting for the final survival classification task. The fusion strategy leveraged feature-level fusion, combining features from different modalities to augment the model's predictive power, thereby utilizing the strengths of each modality and improving the accuracy of breast cancer prognosis prediction.

A hybrid deep learning [65] model was developed to predict molecular subtypes of breast cancer using gene expression data and pathological images. The TCGA-BRCA dataset, consisting of 1098 samples, was used, with 831 samples selected after filtering. Gene expression data was processed using Log2 and CNVs were detected using Affymetrix SNP 6.0. Pathological images were RGB colored. A multimodal fusion framework was constructed, combining gene and image data with feature extraction networks. The fusion process integrated information from both modalities to enhance the prediction of molecular subtypes. The fusion model outperformed both the DNN and CNN models in terms of accuracy and AUC values, demonstrating the effectiveness of the fusion strategy in improving subtype prediction accuracy.

Table 4: Existing research in multi-modal breast cancer diagnosis

Author	Year	Datasets	Fusion strategy	Modality
Sun et al. [62]	2018	METABRIC	Late fusion	Clinical data, Gene expression
Tong et al. [63]	2020	TCGA-BRCA	Encoder-decoder method	Gene Expressions, CNV
Arya and Saha [61]	2021	METABRIC, TCGA-BRCA	Early fusion	Clinical data, Gene expression
Subramanian et al. [64]	2021	TCGA-BRCA	Early fusion	Histopathology images, Clinical data
Liu et al. [65]	2022	TCGA-BRCA	Late fusion	Histopathology images, Gene expressions
Howard et al. [66]	2022	TCGA-BRCA	Late fusion	Histopathology images, Gene expressions
Arya and Saha [67]	2022	METABRIC, TCGA-BRCA	Encoder-decoder method	Clinical data, Gene expression
Arya and Saha [68]	2022	METABRIC	Early fusion	Clinical data, Gene expression
Furtney et al. [69]	2023	TCGA-BRCA	Graph-neural network method	Histopathology images, Clinical data, Gene Expressions, Radiological data
Rani et al. [70]	2023	TCGA-BRCA	Early fusion	Histopathology images, Gene expressions
Kayikci et al. [71]	2023	METABRIC	Attention-based	Clinical data, Gene expression
Arya et al.[72]	2023	TCGA-BRCA	Early fusion	Clinical data, Gene expression
Mondol et al. [73]	2024	TCGA-BRCA	Attention-based	Histopathology images, Clinical data, Gene Expressions
Huang et al. [74]	2024	TCGA-BRCA, GMUCH-BRCA	Early fusion	Histopathology images, Clinical data
Li and Nabavi [75]	2024	TCGA-BRCA	Graph-neural network method	Gene Expressions, CNV

A deep learning model [66] was employed to predict recurrence assay results and the risk of recurrence in breast cancer patients. The dataset utilized was The Cancer Genome Atlas (TCGA), comprising 1,099 slides manually annotated to distinguish tumor from surrounding stroma. Tesselated image tiles were extracted from tumor areas and downscaled using a convolutional neural network backbone with fully connected hidden layers for outcome prediction. The model's predictions were based on a combination of digital histology and clinical risk factors, with patient-level predictions calculated by weighting the average of tile-level predictions according to the likelihood of tumor presence in each tile. This fusion strategy led to improved accuracy in predicting recurrence assay results and risk of recurrence, surpassing traditional clinical nomograms. The application of the fusion strategy facilitated enhanced accuracy in predicting recurrence assay results and risk of recurrence in breast cancer patients.

Canonical Correlation Analysis (CCA) and its penalized variants (pCCA) were employed for multi-modality fusion in breast cancer prediction [64]. CCA identifies

13

correlated linear combinations of two or multiple modalities, while pCCA integrates penalties based on domain knowledge to effectively handle high-dimensional, low-sample-size datasets such as cancer imaging-genomics. The dataset comprised histopathology data and RNA-sequencing data from breast cancer patients in The Cancer Genome Atlas (TCGA). A two-stage prediction pipeline was proposed utilizing pCCA embeddings generated with deflation for latent variable prediction. The fusion strategy amalgamated information from histology and genomics data to augment survival prediction in breast cancer patients. The model demonstrated superior performance compared to Principal Components Analysis (PCA) embeddings in survival prediction tasks. A deep learning approach was proposed for survival risk stratification in breast cancer, integrating histopathological imaging, genetic, and clinical data [73]. The MaxViT model was employed for image feature extraction, with self-attention mechanisms capturing intricate image relationships at the patient level. A dual cross-attention mechanism fused image features with genetic data to enhance predictive accuracy. The study utilized the TCGA-BRCA dataset, consisting of 249 whole-slide images. Clinical variables such as tumor grade, size, patient age, and lymph node status were meticulously prepared for model evaluations. A dual cross-attention mechanism was utilized to refine the interaction between histopathology image features and genetic expression profiles, employing two attention operations: image features as Query and genetic features as Key and Value, and vice versa.

A Multimodal Deep Neural Network (MDNNMD) was proposed for breast cancer prognosis prediction, incorporating multi-dimensional data [62]. The method's architecture design and fusion of multi-dimensional data represented innovative aspects. The METABRIC dataset, obtained from 1,980 breast cancer patients, contained gene expression profiles, CNA profiles, and clinical information. Patients were classified as short-term (491) or long-term survivors based on a 5-year survival threshold, with each patient's data comprising 27 clinical features. The MDNNMD method integrated multi-dimensional data, combining gene expression profiles and CNA profiles for comprehensive analysis. The fusion of multi-dimensional data in MDNNMD outperformed single-dimensional data-based methods such as DNN-Clinical, DNN-Expr, and DNN-CNA, highlighting the advantages of this fusion approach.

A two-stage generative incomplete multi-view prediction model, named GIMPP [67], was introduced to address missing view problems in breast cancer prognosis prediction. The first stage utilized multi-view encoder networks and a bi-modal attention scheme to learn common latent space representations. In the second stage, missing view data was generated using view-specific generative adversarial networks conditioned on shared representations and encoded features from other views. The effectiveness of the method was evaluated on the TCGA-BRCA and METABRIC datasets, commonly utilized in cancer research, to demonstrate its superiority over state-of-the-art methods. Fusion in the model was achieved through the generation of missing view data using view-specific generative adversarial networks. A multimodal siamese model was proposed for breast cancer survival prediction, integrating pathological images and clinical data. Siamese-RegNet was employed to extract survival-related features from pathological patches and capture correlations [74]. The model comprised patch sampling,

feature extraction using Siamese-RegNet, and survival prediction using the LASSO-Cox model. Two datasets, TCGA-BRCA from The Cancer Genome Atlas project and GMUCH-BRCA from Guangzhou Medical University Cancer Hospital, were utilized. The fusion strategy of the model combined information from both images and clinical data to enhance survival prediction accuracy. This comprehensive representation of features facilitated improved survival risk assessment.

A method was utilized to group patient graphs into batches for training and updating graph embeddings in a Graph Neural Network (GNN) framework [69], employing the cross-entropy loss function. The datasets incorporated in the study comprised The Cancer Genome Atlas Breast Invasive Carcinoma (TCGA-BRCA) dataset, encompassing clinical, genomic, and radiological data from 1,040 patients, with a subset of 108 patients featuring DCE-MRI scans and radiologist tumor measurements. The fusion type adopted in the study involved multimodal fusion of breast cancer patient data with graph convolutional neural networks, modeling patient information into graphs using deep learning features extracted from diverse modalities, such as MRI scans and genomic variant assays. The aim was to enhance the model's generalization potential and performance.

4 Explainable breast cancer detection: Challenges, trends and future directions

Explainability remains a critical challenge in breast cancer detection, particularly with the increasing use of complex machine learning and deep learning models. Explainability is essential for clinical decision-making, trust and transparency, regulatory compliance, and error detection and correction. Explainable models enable clinicians to understand the rationale behind a diagnosis, thereby facilitating more informed decision-making and increasing trust in automated systems through transparent conclusions.
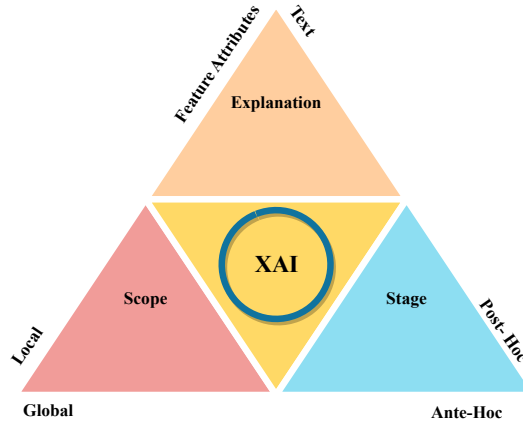

Fig. 5: Different types of XAI methods

However, several challenges in explainability include the complexity of models, data diversity, the black-box nature of algorithms, the trade-off between explainability and accuracy, and the lack of standardization in medical diagnostics. The absence of universally accepted standards for explainability leads to variability in approaches, complicating comparisons and consistent interpretations.

Figure. 5 illustrates how XAI methods can be categorised in different contexts. Based on Explanation, stage and scope, there can be different methods. Exaplaiability explanations can be in terms of feature attributes and textual format. In scope-based categorisation, there are local and global methods. Post-hoc and ante-hoc are the stage-based XAI methods.

To enhance explainability, techniques such as feature importance and saliency maps can provide insights into which aspects of the input are driving the model's predictions. Model-agnostic methods like LIME and SHAP allow for the explanation of any machine learning model, offering flexibility in creating explainable outputs. Interpretable models, such as decision trees or linear models, provide greater transparency, albeit potentially at the cost of reduced accuracy. Additionally, human feedback and oversight in the diagnostic process can help validate and explain automated decisions, combining algorithmic efficiency with human intuition. In conclusion, explainability is crucial in breast cancer detection to ensure reliable and trustworthy outcomes, requiring a combination of technical solutions, regulatory compliance, and human oversight to create models that are both accurate and transparent.

4.1 An explainability analysis on multi-modal techniques

In the domain of uni-modal breast cancer detection, significant advancements have been made in integrating explainable artificial intelligence (AI) techniques to enhance the interpretability and reliability of predictive models. Gu et al [76] developed an auxiliary decision support system that combines ensemble learning with case-based reasoning (CBR) to predict breast cancer recurrence. Using XGBoost for predictions and CBR for providing comprehensible explanations, this system effectively communicated the importance of various attributes, aligning well with human reasoning and gaining acceptance among clinicians. Kabakçı et al. [77] proposed an automated method for determining CerbB2/HER2 scores from breast tissue images by adhering to ASCO/CAP recommendations. This method employed cell-based image analysis and a hand-crafted feature extraction approach, ensuring both interpretability and adaptability to guideline updates without the need for re-training.

Moreover, recent studies have focused on enhancing the explainability of deep learning models used in breast cancer histopathology. Authors of [78] utilized pre-trained models combined with gradient-boosting classifiers to achieve high accuracy in classifying breast cancer images from the BreakHis dataset. Similarly, Peta et al. [79] introduced an explainable deep learning technique involving adaptive unsharp mask filtering and the Explainable Soft Attentive EfficientNet (ESAE-Net), which provided improved visualization and understanding of classification decisions. Jaume et al. [80] presented CGEXPLAINER, a post-hoc explainer for graph representations in digital pathology, which pruned redundant graph components to maximize mutual

information between the original prediction and the sub-graph explanation. These contributions, along with methods like the cost-sensitive CatBoost classifier with LIME explainer [81] and the use of SHAP for feature importance analysis in tumour cellularity assessment [82], highlight the growing emphasis on explainability to ensure that AI systems for breast cancer detection are not only accurate but also interpretable and trustworthy for clinical application.

Explainability is a critical factor in radio genomics [83], as it fosters trust with end-users like physicians and patients, driving the deployment of deep learning models in research and clinical practice. It increases confidence in the model's decision-making process, enabling better understanding and acceptance of results. Explainability also serves as a debugging process for model training and fine-tuning, identifying potential errors or biases. It also helps bypass malicious manipulation, ensuring the integrity and security of radiogenomic research and its applications. In the healthcare field, explainability is especially important as it facilitates better interpretation and understanding of complex AI models, leading to improved patient care and treatment outcomes.

Holzinger et al. [84] proposed the utilization of Graph Neural Networks (GNNs) as a method for achieving multi-modal causability within explainable AI (xAI). This approach facilitated information fusion through the establishment of causal links between features using graph structures. The method's objective was to construct a multi-modal feature representation space, utilizing knowledge bases as initial connectors for the development of novel explanation interface techniques. Essential components included intra-modal feature extraction and multi-modal embedding. Various GNN architectures and graph embeddings, such as GCNN, Graph Isomorphism Network (GIN), and SchNet, were considered viable options. Additionally, dynamic GNN architectures like Pointer Graph Networks (PGN) were employed to enable the processing of adaptive graphs. Zhang et al. [85] introduced a Deep Multimodal Reasoning and Fusion Network (DMRFNet) for Visual Question Answering (VQA) and explanation generation. The model employed multimodal reasoning and fusion techniques to improve the accuracy of answers and explanations. A key innovation was the Multi-Graph Reasoning and Fusion (MGRF) layer, which utilized pre-trained semantic relation embeddings to handle complex spatial and semantic relations among visual objects. DMRFNet was capable of being stacked in depth to facilitate comprehensive reasoning and fusion of multimodal relations. Additionally, an explanation generation module was incorporated to provide justifications for predicted answers. Experimental findings demonstrated the model's effectiveness in achieving both quantitative and qualitative performance improvements.

Authors of [86] introduced a segmentation framework with an interpretation module that highlights critical features from each modality, guided by a novel interpretation loss with strengthened and perturbed fusion schemes. This approach effectively generates meaningful interpretable masks, improving multi-modality information integration and segmentation performance. Visualization and perturbation experiments validate the effectiveness of the interpretation method in exploiting meaningful features from each modality. An interpretable decision-support model for breast cancer diagnosis using histopathology images was proposed in [87]. This method integrated

an attention branch into a variant of the DarkNet19 CNN model to enhance interpretability and performance. The attention branch generated a heatmap to identify regions of interest, while a perception branch performed image classification through a fully connected layer. Training and validation utilized over 7,000 breast cancer biopsy slide images from the BreaKHis dataset, resulting in a binary classification accuracy of 98.7%. Notably, the model offered enhanced clinical interpretability, with highlighted cancer regions corresponding well with expert pathologist findings. The ABN-DCN model effectively combined an attention mechanism with a CNN feature extractor, thereby improving both diagnostic interpretability and classification performance in histopathology images.

5 New frontiers in explainability and multi-modality

Recent advancements in multimodality and explainability in medical diagnostics can bring significant improvements in breast cancer detection. These developments highlight potential future directions for integrating advanced computational models and explainability methods into histopathology-based breast cancer diagnostics. In the context of multimodal methodologies, multimodal fusion is a subset of techniques within the broader field of multimodal analysis. It should be noted that multimodal fusion is prevalent, but that other approaches exist within this field as well. As part of a contemporary focus in multimodal methodologies, image and textual data are integrated, manifesting in applications including report generation [88], Visual Question Answering (VQA) [89], cross-modal retrieval [90], and semantic segmentation [91]. It has been noted that substantial scholarly attention has been devoted to the leveraging of medical image and text data through these methodologies[1]. Nevertheless, there still remains a need for further investigation, especially in the context of predicating the diagnosis of breast cancer based on histopathological data using multimodal approaches.

X-VARS [92], a multimodal large language model initially designed for football refereeing tasks, utilized Video-ChatGPT to process video features and predict responses. This model emphasized interpretability and has demonstrated strong performance in human studies, indicating its potential for adaptation in breast cancer detection. By integrating diverse data sources, such as histopathology images and clinical records, similar models could offer comprehensible diagnostic support, thereby enhancing the accuracy and transparency of the diagnostic process. The LeGrad [93] explainability method, which employs Vision Transformers (ViTs) [94], utilizes techniques such as GradCAM [95] and AttentionCAM [96] to provide granular insights into feature formation. These explainability methods are crucial for breast cancer detection, offering transparent interpretations of model decisions. By adapting this method to a multimodal scenario that includes histopathology images and clinical/textual data, it can provide comprehensive diagnostic support. This integration enhances trust and clinical applicability by offering transparent and interpretable insights across various data types, thereby improving the accuracy and reliability of breast cancer diagnostics.

The method proposed by Hu et al. [90] for fine-grained cross-modal alignment between histopathology WSIs and diagnostic reports holds promise as a future avenue

in explainable multimodal breast cancer detection. By leveraging anchor-based WSI and prompt-based text encoders, this method ensured that relevant diagnostic information was accessible and interpretable to pathologists. Through precise alignment and interpretation of multimodal diagnostic data, including histopathology images and clinical and textual reports, the method enhances transparency and interpretability in breast cancer diagnosis. This approach can provide clear insights into the decision-making process of diagnostic models, thereby enhancing trust and clinical acceptance in the application of multimodal AI systems for breast cancer detection. A multimodal image search strategy was described in [97] as a method of improving diagnosis, prognosis, and prediction in histopathology. With this method, large image archives can be explored to identify patterns and correlations using foundation models for feature extraction and image matching. A breast cancer detection framework based on this framework could provide efficient retrieval and comparison of histopathological images, thereby aiding in the identification of malignancies and their characteristics.

Investigating local surrogate explainability techniques in deep learning models, researchers explored the use of VisualBERT and UNITER networks to generate multimodal visual and language explanations [98]. The potential of these models to mimic domain expertise underscores the value of explainable AI techniques in breast cancer detection. By providing clear and understandable rationales for automated decisions, such methods enhance clinical trust and support informed decision-making in diagnostic processes. A framework named LangXAI [99] was introduced, integrating explainable AI with advanced vision models to generate textual explanations for visual recognition tasks. This framework enhances transparency and plausibility, potentially improving breast cancer detection by making the diagnostic process more understandable and reliable for clinicians. Consequently, it supports better patient outcomes.

Various explainable AI methods, including Gradient backpropagation and Integrated-Gradients, were applied in [100] to analyze the MedCLIP model. These methods provided valuable insights into model predictions, offering pivotal information for the development of breast cancer detection models. Ensuring the transparency and comprehensibility of model decisions can play a crucial role in facilitating regulatory compliance and fostering clinical acceptance of such models in diagnostic settings. A tool called LVLM-Interpret [101] was developed to interpret responses from large vision-language models, employing techniques such as raw attention and relevancy maps. This tool's capacity to visualize and comprehend model outputs can be utilized in breast cancer detection to improve the interpretability and reliability of AI-driven diagnostic tools.

An ex-ILP [102] framework was introduced to enhance reasoning capabilities in vision-language models by Yang et al. (2023). By improving implicit reasoning skills, this methodology could be harnessed in breast cancer detection to interpret complex interactions between visual and textual data, thus contributing to more accurate and nuanced diagnostic insights. The NLX-GPT method, introduced in [103], integrated discriminative answer prediction and explanation tasks into a unified model. This approach, which achieves high performance across diverse tasks, holds the potential for adaptation in breast cancer detection. By providing both diagnostic conclusions and

19

their explanations, the NLX-GPT method enhances the usability and trustworthiness of AI models in clinical settings.

These advancements indicate a promising future for the integration of multimodal fusion and explainability in breast cancer diagnostics. By leveraging these innovative approaches, histopathology-based breast cancer diagnosis can be significantly enhanced, leading to improved accuracy, transparency, and ultimately, patient outcomes.

5.1 Unified Framework: Leveraging Multi-Modal and Explainability

The framework proposed for multimodal explainable breast cancer diagnosis involves a systematic process aimed at enhancing diagnostic accuracy and transparency while integrating human expertise for improved patient outcomes. Figure. 6 illustrates the proposed framework. In the initial step, histopathology images are processed using pre-trained medical report generation models. These models, such as CLARA [104], automatically generate comprehensive reports from the images, augmenting them with relevant features extracted through computer vision techniques. Clinical data, including patient history and laboratory results, are integrated into the report generation pipeline to ensure contextually relevant diagnostic reports.
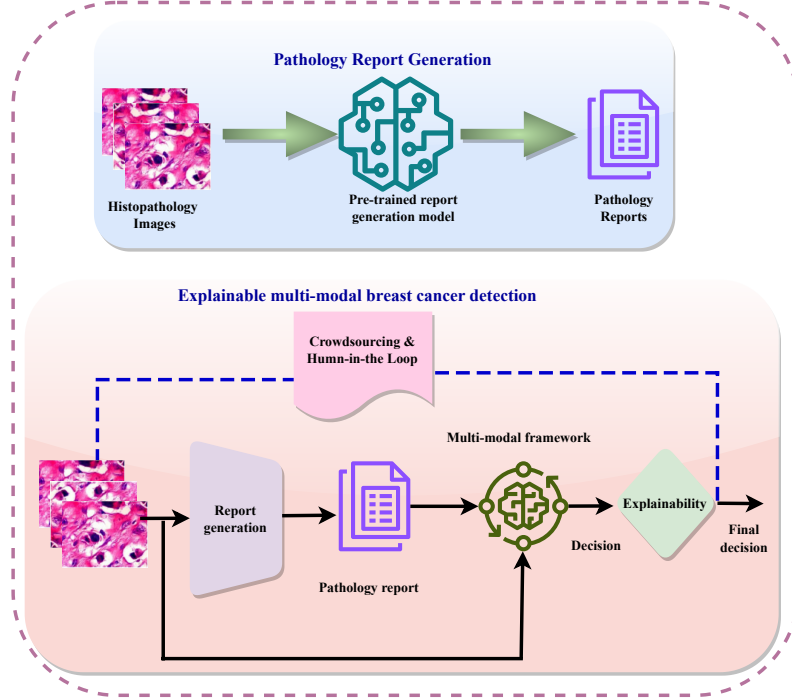

Fig. 6: Proposed framework for explainable multi-modal breast cancer detection

Subsequently, in the multimodal explainable framework for diagnosis, the generated diagnostic reports and histopathology images serve as input. Visual language models, such as Vision Transformers or large language models, are employed to process both visual and textual information simultaneously. Explainability techniques like GradCAM and AttentionCAM are implemented to provide interpretable insights into model decisions, enhancing transparency and trust in the diagnostic process. Model outputs are visualized using tools like LVLM-Interpret to improve interpretability and reliability.

Human expertise is integrated through crowdsourcing or expert consultations to validate and refine model predictions, ensuring clinical relevance and accuracy. This human-in-the-loop approach facilitates informed decision-making and iterative refinement based on feedback from clinicians and patients. Ultimately, the framework supports diagnostic support by providing transparent and understandable diagnostic conclusions, along with explanations for model predictions. It is integrated into existing clinical workflows to streamline diagnostic processes and enhance patient care, contributing to advancements in the field of breast cancer diagnostics.

6 Conclusion

Detection of breast cancer at an early stage and with accuracy is essential to improving patient outcomes. Despite the fact that traditional diagnostic methods primarily rely on single-modal approaches, data analytics combined with diverse sources of data represent a significant shift in diagnostic paradigms. A major step forward in breast cancer diagnosis has been made possible through the adoption of multimodal techniques, which combine both image and non-image data. This review has delved into the burgeoning field of multimodal methodologies, with a specific focus on the fusion of histopathology images with non-image data. Furthermore, the incorporation of Explainable AI (XAI) serves to illuminate the decision-making processes of intricate algorithms, underlining the importance of transparency in diagnostic procedures. By leveraging multi-modal data and emphasizing explainability, this review advocates for enhancing diagnostic accuracy, bolstering clinician confidence, and fostering patient engagement. Moreover, these advancements aim to facilitate the development of personalized treatment strategies tailored to the unique needs of each patient. In addition to addressing the current landscape of multimodality and explainability, this review identifies pertinent research gaps, thereby guiding the trajectory of future studies in this field. By contributing to a comprehensive understanding of multi-modal diagnostic techniques and the imperative of explainability, this review seeks to inform strategic directions in breast cancer diagnosis and treatment, ultimately striving for improved patient outcomes and a more effective healthcare landscape.

Acknowledgment

References

[1] Sun, Z., Lin, M., Zhu, Q., Xie, Q., Wang, F., Lu, Z., Peng, Y.: A scoping review on multimodal deep learning in biomedical images and texts. Journal of Biomedical Informatics, 104482 (2023)

[2] Krithiga, R., Geetha, P.: Breast cancer detection, segmentation and classification on histopathology images analysis: a systematic review. Archives of Computational Methods in Engineering **28**(4), 2607–2619 (2021)

[3] Tafavvoghi, M., Bongo, L.A., Shvetsov, N., Busund, L.-T.R., Møllersen, K.: Publicly available datasets of breast histopathology h&e whole-slide images: A scoping review. Journal of Pathology Informatics, 100363 (2024)

[4] Abo-El-Rejal, A., Ayman, S., Aymen, F.: Advances in breast cancer segmentation: A comprehensive review. Acadlore Transactions on AI and Machine Learning **3**(2), 70–83 (2024)

[5] Hussain, S., Ali, M., Naseem, U., Nezhadmoghadam, F., Jatoi, M.A., Gulliver, T.A., Tamez-Peña, J.G.: Breast cancer risk prediction using machine learning: a systematic review. Frontiers in Oncology **14**, 1343627 (2024)

[6] Brodhead, M., Woods, R.W., Fowler, A.M., Roy, M., Neuman, H., Gegios, A.: Multimodality imaging review of metastatic melanoma involving the breast. Current Problems in Diagnostic Radiology (2024)

[7] Luo, L., Wang, X., Lin, Y., Ma, X., Tan, A., Chan, R., Vardhanabhuti, V., Chu, W.C., Cheng, K.-T., Chen, H.: Deep learning in breast cancer imaging: A decade of progress and future directions. IEEE Reviews in Biomedical Engineering (2024)

[8] Rautela, K., Kumar, D., Kumar, V.: A comprehensive review on computational techniques for breast cancer: past, present, and future. Multimedia Tools and Applications, 1–34 (2024)

[9] Singh, A., Kaur, S., Singh, D., Singh, G.: Technical review of breast cancer screening and detection using artificial intelligence and radiomics. In: 2024 11th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 1171–1176 (2024). IEEE

[10] Thakur, N., Kumar, P., Kumar, A.: A systematic review of machine and deep learning techniques for the identification and classification of breast cancer through medical image modalities. Multimedia Tools and Applications **83**(12), 35849–35942 (2024)

[11] Nasser, M., Yusof, U.K.: Deep learning based methods for breast cancer diagnosis: a systematic review and future direction. Diagnostics **13**(1), 161

(2023)

[12] Obeagu, E.I., Obeagu, G.U.: Breast cancer: A review of risk factors and diagnosis. Medicine **103**(3), 36905 (2024)

[13] Rai, H.M.: Cancer detection and segmentation using machine learning and deep learning techniques: A review. Multimedia Tools and Applications **83**(9), 27001–27035 (2024)

[14] Liu, Z., Lin, F., Huang, J., Wu, X., Wen, J., Wang, M., Ren, Y., Wei, X., Song, X., Qin, J., *et al.*: A classifier-combined method for grading breast cancer based on dempster-shafer evidence theory. Quantitative Imaging in Medicine and Surgery **13**(5), 3288 (2023)

[15] Kumaraswamy, E., Kumar, S., Sharma, M.: An invasive ductal carcinomas breast cancer grade classification using an ensemble of convolutional neural networks. Diagnostics **13**(11), 1977 (2023)

[16] Huang, Y., Zeng, P., Zhong, C.: Classifying breast cancer subtypes on multi-omics data via sparse canonical correlation analysis and deep learning. BMC bioinformatics **25**(1), 132 (2024)

[17] Choi, J.M., Chae, H.: mobrca-net: a breast cancer subtype classification framework based on multi-omics attention neural networks. BMC bioinformatics **24**(1), 169 (2023)

[18] Raza, A., Ullah, N., Khan, J.A., Assam, M., Guzzo, A., Aljuaid, H.: Deep-breastcancernet: A novel deep learning model for breast cancer detection using ultrasound images. Applied Sciences **13**(4), 2082 (2023)

[19] Al-Dhabyani, W., Gomaa, M., Khaled, H., Fahmy, A.: Dataset of breast ultrasound images. Data in brief **28**, 104863 (2020)

[20] Paulo, S.: Breast ultrasound image. Mendeley data (2017)

[21] The Cancer Genome Atlas (TCGA): Genomic Data Commons Data Portal (GDC). https://portal.gdc.cancer.gov/projects/TCGA-BRCA. Accessed 07 Jul. 2023

[22] Parshionikar, S., Bhattacharyya, D.: An enhanced multi-scale deep convolutional orchard capsule neural network for multi-modal breast cancer detection. Healthcare Analytics **5**, 100298 (2024)

[23] Spanhol, F.A., Oliveira, L.S., Petitjean, C., Heutte, L.: A dataset for breast cancer histopathological image classification. Ieee transactions on biomedical engineering **63**(7), 1455–1462 (2015)

[24] Zuluaga-Gomez, J., Al Masry, Z., Benaggoune, K., Meraghni, S., Zerhouni, N.:

A cnn-based methodology for breast cancer diagnosis using thermal images. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization **9**(2), 131–145 (2021)

[25] DataBioX: DataBioX Datasets. Accessed: 2024-06-02 (2024). https://databiox. com/datasets/

[26] Sahu, Y., Tripathi, A., Gupta, R.K., Gautam, P., Pateriya, R.K., Gupta, A.: A cnn-svm based computer aided diagnosis of breast cancer using histogram k-means segmentation technique. Multimedia Tools and Applications **82**(9), 14055–14075 (2023)

[27] Sahu, A., Das, P.K., Meher, S.: High accuracy hybrid cnn classifiers for breast cancer detection using mammogram and ultrasound datasets. Biomedical Signal Processing and Control **80**, 104292 (2023)

[28] Lekamlage, C.D., Afzal, F., Westerberg, E., Cheddad, A.: Mini-ddsm: Mammography-based automatic age estimation. In: 2020 3rd International Conference on Digital Medicine and Image Processing, pp. 1–6 (2020)

[29] Srikantamurthy, M.M., Rallabandi, V.S., Dudekula, D.B., Natarajan, S., Park, J.: Classification of benign and malignant subtypes of breast cancer histopathology imaging using hybrid cnn-lstm based transfer learning. BMC Medical Imaging **23**(1), 19 (2023)

[30] Guo, D., Lu, C., Chen, D., Yuan, J., Duan, Q., Xue, Z., Liu, S., Huang, Y.: A multimodal breast cancer diagnosis method based on knowledge-augmented deep learning. Biomedical Signal Processing and Control **90**, 105843 (2024)

[31] Liu, H., Shi, Y., Li, A., Wang, M.: Multi-modal fusion network with intra-and inter-modality attention for prognosis prediction in breast cancer. Computers in Biology and Medicine **168**, 107796 (2024)

[32] Sivamurugan, J., Sureshkumar, G.: Applying dual models on optimized lstm with u-net segmentation for breast cancer diagnosis using mammogram images. Artificial Intelligence in Medicine **143**, 102626 (2023)

[33] Kendall, E.J., Barnett, M.G., Chytyk-Praznik, K.: Automatic detection of anomalies in screening mammograms. BMC Medical Imaging **13**, 1–11 (2013)

[34] Murata, T., Yoshida, M., Shiino, S., Ogawa, A., Watase, C., Satomi, K., Jimbo, K., Maeshima, A., Iwamoto, E., Takayama, S., *et al.*: A prediction model for distant metastasis after isolated locoregional recurrence of breast cancer. Breast Cancer Research and Treatment **199**(1), 57–66 (2023)

[35] Hussein, M., Elnahas, M., Keshk, A.: A framework for predicting breast cancer recurrence. Expert Systems with Applications **240**, 122641 (2024)

[36] Ahmed, L., Iqbal, M.M., Aldabbas, H., Khalid, S., Saleem, Y., Saeed, S.: Images data practices for semantic segmentation of breast cancer using deep neural network. Journal of Ambient Intelligence and Humanized Computing **14**(11), 15227–15243 (2023)

[37] Lee, R.S., Gimenez, F., Hoogi, A., Miyake, K.K., Gorovoy, M., Rubin, D.L.: A curated mammography data set for use in computer-aided detection and diagnosis research. Scientific data **4**(1), 1–9 (2017)

[38] Alam, T., Shia, W.-C., Hsu, F.-R., Hassan, T.: Improving breast cancer detection and diagnosis through semantic segmentation using the unet3+ deep learning framework. Biomedicines **11**(6), 1536 (2023)

[39] Prinzi, F., Insalaco, M., Orlando, A., Gaglio, S., Vitabile, S.: A yolo-based model for breast cancer detection in mammograms. Cognitive Computation **16**(1), 107–120 (2024)

[40] Moreira, I.C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M.J., Cardoso, J.S.: Inbreast: toward a full-field digital mammographic database. Academic radiology **19**(2), 236–248 (2012)

[41] Guo, H., Li, M., Liu, H., Chen, X., Cheng, Z., Li, X., Yu, H., He, Q.: Multi-threshold image segmentation based on an improved salp swarm algorithm: Case study of breast cancer pathology images. Computers in Biology and Medicine **168**, 107769 (2024)

[42] Rajoub, B., Qusa, H., Abdul-Rahman, H., Mohamed, H.: Segmentation of breast tissue structures in mammographic images. Artificial Intelligence and Image Processing in Medical Imaging, 115–146 (2024)

[43] Soliman, A., Li, Z., Parwani, A.V.: Artificial intelligence's impact on breast cancer pathology: a literature review. Diagnostic Pathology **19**(1), 1–18 (2024)

[44] Gallagher, W.M., McCaffrey, C., Jahangir, C., Murphy, C., Burke, C., Rahman, A.: Artificial intelligence in digital histopathology for predicting patient prognosis and treatment efficacy in breast cancer. Expert Review of Molecular Diagnostics (just-accepted) (2024)

[45] Sweetlin, E.J., Saudia, S.: A review of machine learning algorithms on different breast cancer datasets. In: International Conference on Big Data, Machine Learning, and Applications, pp. 659–673 (2021). Springer

[46] Heiliger, L., Sekuboyina, A., Menze, B., Egger, J., Kleesiek, J.: Beyond medical imaging-a review of multimodal deep learning in radiology. Authorea Preprints (2023)

[47] Laokulrath, N., Gudi, M.A., Deb, R., Ellis, I.O., Tan, P.H.: Invasive breast

cancer reporting guidelines: Iccr, cap, rcpath, rcpa datasets and future directions. Diagnostic Histopathology (2024)

[48] Brancati, N., Anniciello, A.M., Pati, P., Riccio, D., Scognamiglio, G., Jaume, G., De Pietro, G., Di Bonito, M., Foncubierta, A., Botti, G., *et al.*: Bracs: A dataset for breast carcinoma subtyping in h&e histology images. Database **2022**, 093 (2022)

[49] Aksac, A., Demetrick, D.J., Ozyer, T., Alhajj, R.: Brecahad: a dataset for breast cancer histopathological annotation and diagnosis. BMC research notes **12**, 1–3 (2019)

[50] Martel, A., Nofech-Mozes, S., Salama, S., Akbar, S., Peikari, M.: Assessment of residual breast cancer cellularity after neoadjuvant chemotherapy using digital pathology [data set]. The Cancer Imaging Archive (2019)

[51] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium: The Clinical Proteomic Tumor Analysis Consortium Breast Invasive Carcinoma Collection (CPTAC-BRCA). The Cancer Imaging Archive. Accessed 07 Jul. 2023 (2020). https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=70227748

[52] Yan, R., Zhang, F., Rao, X., Lv, Z., Li, J., Zhang, L., Liang, S., Li, Y., Ren, F., Zheng, C., *et al.*: Richer fusion network for breast cancer classification based on multimodal data. BMC Medical Informatics and Decision Making **21**, 1–15 (2021)

[53] vEarly Breast Cancer Core-Needle Biopsy WSI (BCNB): Grand Challenge. https://bcnb.grand-challenge.org/. Accessed 07 Jul. 2023 (2022)

[54] Huang, Z., Shao, W., Han, Z., Alkashash, A.M., Sancha, C., Parwani, A.V., Nitta, H., Hou, Y., Wang, T., Salama, P., *et al.*: Artificial intelligence reveals features associated with breast cancer neoadjuvant chemotherapy responses from multi-stain histopathologic images. NPJ Precision Oncology **7**(1), 14 (2023)

[55] The Genotype-Tissue Expression (GTEx): GTEx Portal. https://gtexportal.org/home/histologyPage. Accessed 07 Jul. 2023

[56] Lipkova, J., Chen, R.J., Chen, B., Lu, M.Y., Barbieri, M., Shao, D., Vaidya, A.J., Chen, C., Zhuang, L., Williamson, D.F., *et al.*: Artificial intelligence for multimodal data integration in oncology. Cancer cell **40**(10), 1095–1110 (2022)

[57] Zhao, F., Zhang, C., Geng, B.: Deep multimodal data fusion. ACM Computing Surveys (2024)

[58] Yellapragada, S., Graikos, A., Prasanna, P., Kurc, T., Saltz, J., Samaras, D.:

Pathldm: Text conditioned latent diffusion model for histopathology. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 5182–5191 (2024)

[59] Ding, R., Hall, J., Tenenholtz, N., Severson, K.: Improving mitosis detection on histopathology images using large vision-language models. arXiv preprint arXiv:2310.07176 (2023)

[60] Lu, M.Y., Chen, B., Zhang, A., Williamson, D.F., Chen, R.J., Ding, T., Le, L.P., Chuang, Y.-S., Mahmood, F.: Visual language pretrained multiple instance zero-shot transfer for histopathology images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 19764–19775 (2023)

[61] Arya, N., Saha, S.: Multi-modal advanced deep learning architectures for breast cancer survival prediction. Knowledge-Based Systems **221**, 106965 (2021)

[62] Sun, D., Wang, M., Li, A.: A multimodal deep neural network for human breast cancer prognosis prediction by integrating multi-dimensional data. IEEE/ACM transactions on computational biology and bioinformatics **16**(3), 841–850 (2018)

[63] Tong, L., Mitchel, J., Chatlin, K., Wang, M.D.: Deep learning based feature-level integration of multi-omics data for breast cancer patients survival analysis. BMC medical informatics and decision making **20**, 1–12 (2020)

[64] Subramanian, V., Syeda-Mahmood, T., Do, M.N.: Multi-modality fusion using canonical correlation analysis methods: Application in breast cancer survival prediction from histology and genomics. arXiv preprint arXiv:2111.13987 (2021)

[65] Liu, T., Huang, J., Liao, T., Pu, R., Liu, S., Peng, Y.: A hybrid deep learning model for predicting molecular subtypes of human breast cancer using multimodal data. Irbm **43**(1), 62–74 (2022)

[66] Howard, F.M., Dolezal, J., Kochanny, S., Khramtsova, G., Vickery, J., Srisuwananukorn, A., Woodard, A., Chen, N., Nanda, R., Perou, C.M., et al.: Multimodal prediction of breast cancer recurrence assays and risk of recurrence. bioRxiv, 2022–07 (2022)

[67] Arya, N., Saha, S.: Generative incomplete multi-view prognosis predictor for breast cancer: Gimpp. IEEE/ACM Transactions on Computational Biology and Bioinformatics **19**(4), 2252–2263 (2021)

[68] Arya, N., Saha, S.: Multi-modal classification for human breast cancer prognosis prediction: proposal of deep-learning based stacked ensemble model. IEEE/ACM transactions on computational biology and bioinformatics **19**(2), 1032–1041 (2020)

[69] Furtney, I., Bradley, R., Kabuka, M.R.: Patient graph deep learning to predict breast cancer molecular subtype. IEEE/ACM transactions on computational biology and bioinformatics (2023)

[70] Rani, S., Ahmad, T., Masood, S., Saxena, C.: Diagnosis of breast cancer molecular subtypes using machine learning models on unimodal and multimodal datasets. Neural Computing and Applications **35**(34), 24109–24121 (2023)

[71] Kayikci, S., Khoshgoftaar, T.M.: Breast cancer prediction using gated attentive multimodal deep learning. Journal of Big Data **10**(1), 62 (2023)

[72] Arya, N., Saha, S., Mathur, A., Saha, S.: Improving the robustness and stability of a machine learning model for breast cancer prognosis through the use of multi-modal classifiers. Scientific Reports **13**(1), 4079 (2023)

[73] Mondol, R.K., Millar, E.K., Sowmya, A., Meijering, E.: Mm-survnet: Deep learning-based survival risk stratification in breast cancer through multimodal data fusion. arXiv preprint arXiv:2402.11788 (2024)

[74] Huang, S., Liu, Z., Liu, Z.: Multimodal siamese model for breast cancer survival prediction. In: 2024 4th International Conference on Neural Networks, Information and Communication (NNICE), pp. 925–928 (2024). IEEE

[75] Li, B., Nabavi, S.: A multimodal graph neural network framework for cancer molecular subtype classification. BMC bioinformatics **25**(1), 27 (2024)

[76] Gu, D., Su, K., Zhao, H.: A case-based ensemble learning system for explainable breast cancer recurrence prediction. Artificial Intelligence in Medicine **107**, 101858 (2020)

[77] Kabakçı, K.A., Çakır, A., Türkmen, İ., Töreyin, B.U., Çapar, A.: Automated scoring of cerbb2/her2 receptors using histogram based analysis of immunohistochemistry breast cancer tissue images. Biomedical Signal Processing and Control **69**, 102924 (2021)

[78] Maleki, A., Raahemi, M., Nasiri, H.: Breast cancer diagnosis from histopathology images using deep neural network and xgboost. Biomedical Signal Processing and Control **86**, 105152 (2023)

[79] Peta, J., Koppu, S.: Explainable soft attentive efficientnet for breast cancer classification in histopathological images. Biomedical Signal Processing and Control **90**, 105828 (2024)

[80] Jaume, G., Pati, P., Foncubierta-Rodriguez, A., Feroce, F., Scognamiglio, G., Anniciello, A.M., Thiran, J.-P., Goksel, O., Gabrani, M.: Towards explainable graph representations in digital pathology. arXiv preprint arXiv:2007.00311 (2020)

[81] Maouche, I., Terrissa, L.S., Benmohammed, K., Zerhouni, N.: An explainable ai approach for breast cancer metastasis prediction based on clinicopathological data. IEEE Transactions on Biomedical Engineering (2023)

[82] Altini, N., Puro, E., Taccogna, M.G., Marino, F., De Summa, S., Saponaro, C., Mattioli, E., Zito, F.A., Bevilacqua, V.: Tumor cellularity assessment of breast histopathological slides via instance segmentation and pathomic features explainability. Bioengineering **10**(4), 396 (2023)

[83] Liu, Q., Hu, P.: Extendable and explainable deep learning for pan-cancer radiogenomics research. Current opinion in chemical biology **66**, 102111 (2022)

[84] Holzinger, A., Malle, B., Saranti, A., Pfeifer, B.: Towards multi-modal causability with graph neural networks enabling information fusion for explainable ai. Information Fusion **71**, 28–37 (2021)

[85] Zhang, W., Yu, J., Zhao, W., Ran, C.: Dmrfnet: deep multimodal reasoning and fusion for visual question answering and explanation generation. Information Fusion **72**, 70–79 (2021)

[86] Kang, S., Chen, Z., Li, L., Lu, W., Qi, X.S., Tan, S.: Learning feature fusion via an interpretation method for tumor segmentation on pet/ct. Applied Soft Computing **148**, 110825 (2023)

[87] Krishna, S., Suganthi, S., Bhavsar, A., Yesodharan, J., Krishnamoorthy, S.: An interpretable decision-support model for breast cancer diagnosis using histopathology images. Journal of Pathology Informatics **14**, 100319 (2023)

[88] Guo, Z., Ma, J., Xu, Y., Wang, Y., Wang, L., Chen, H.: Histgen: Histopathology report generation via local-global feature encoding and cross-modal context interaction. arXiv preprint arXiv:2403.05396 (2024)

[89] Hartsock, I., Rasool, G.: Vision-language models for medical report generation and visual question answering: A review. arXiv preprint arXiv:2403.02469 (2024)

[90] Hu, D., Jiang, Z., Shi, J., Xie, F., Wu, K., Tang, K., Cao, M., Huai, J., Zheng, Y.: Histopathology language-image representation learning for fine-grained digital pathology cross-modal retrieval. Medical Image Analysis **95**, 103163 (2024)

[91] Van Rijthoven, M., Balkenhol, M., Siliņa, K., Van Der Laak, J., Ciompi, F.: Hooknet: Multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images. Medical image analysis **68**, 101890 (2021)

[92] Held, J., Itani, H., Cioppa, A., Giancola, S., Ghanem, B., Van Droogenbroeck, M.: X-vars: Introducing explainability in football refereeing with multi-modal

large language model. arXiv preprint arXiv:2404.06332 (2024)

[93] Bousselham, W., Boggust, A., Chaybouti, S., Strobelt, H., Kuehne, H.: Legrad: An explainability method for vision transformers via feature formation sensitivity. arXiv preprint arXiv:2404.03214 (2024)

[94] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)

[95] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)

[96] Chen, H., Zhang, X., Guo, Z., Ying, N., Yang, M., Guo, C.: Actnet: Attention based cnn and transformer network for respiratory rate estimation. Biomedical Signal Processing and Control **96**, 106497 (2024)

[97] Tizhoosh, H., Pantanowitz, L.: On image search in histopathology. Journal of Pathology Informatics, 100375 (2024)

[98] Werner, P., Zapaishchykova, A., Ratan, U.: The ability of image-language explainable models to resemble domain expertise. arXiv preprint arXiv:2209.09310 (2022)

[99] Nguyen, T.T.H., Clement, T., Nguyen, P.T.L., Kemmerzell, N., Truong, V.B., Nguyen, V.T.K., Abdelaal, M., Cao, H.: Langxai: Integrating large vision models for generating textual explanations to enhance explainability in visual perception tasks. arXiv preprint arXiv:2402.12525 (2024)

[100] Rehman Hashmi, A.U., Mahapatra, D., Yaqub, M.: Envisioning medclip: A deep dive into explainability for medical vision-language models. arXiv e-prints, 2403 (2024)

[101] Ben Melech Stan, G., Yehezkel Rohekar, R., Gurwicz, Y., Olson, M.L., Bhiwandi-walla, A., Aflalo, E., Wu, C., Duan, N., Tseng, S.-Y., Lal, V.: Lvlm-intrepret: An interpretability tool for large vision-language models. arXiv e-prints, 2404 (2024)

[102] Yang, X., Liu, F., Lin, G.: Neural logic vision language explainer. IEEE Transactions on Multimedia (2023)

[103] Sammani, F., Deligiannis, N.: Uni-nlx: Unifying textual explanations for vision and vision-language tasks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4634–4639 (2023)

[104] Biswal, S., Xiao, C., Glass, L.M., Westover, B., Sun, J.: Clara: clinical report auto-completion. In: Proceedings of The Web Conference 2020, pp. 541–550 (2020)