Multi-Modal Data Integration in Cancer Detection: A Literature Review

Julia Kurnaeva

March 12, 2025

I Introduction

Cancer detection is a critical area where advancements in artificial intelligence (AI) and machine learning (ML) have shown significant promise. However, despite the progress made, further improvements in the accuracy and reliability of these models remain highly desirable [1]. Currently, AI/ML models often rely on single data modalities, such as imaging or genomic data, that can limit the performance in complex cancer detection scenarios.

Recent studies have highlighted the effectiveness of AI/ML models in cancer detection using single modalities [1], [2], [3]. For instance, deep learning techniques have been shown to excel in medical image classification across various modalities, achieving high sensitivity and efficiency in feature extraction compared to traditional methods [1]. Additionally, systematic reviews have underscored the rapid growth and diverse applications of deep learning-based object detection in medical imaging, although these studies primarily focus on single-modality data [2]. Furthermore, research has demonstrated the potential of integrating clinical and genomic data to improve cancer stage classification [2], [3], but studies such as [1], [2], [3] often do not address the broader scope of multi-modal integration for cancer detection .

A significant research gap exists in understanding how multi-modal data integration impacts the accuracy of AI/ML models in detecting cancers such as lung, pancreatic, breast, skin, and osteosarcoma. This gap is evident from the limited focus on multi-modal integration in existing literature [1], [2], [3].

The research question is: **How does multi-modal data integration impact the accuracy of AI/ML models in detecting lung, pancreatic, breast, skin, and osteosarcoma cancers?** Exploring this question can provide a deeper understanding of how integrating multiple data modalities can improve the diagnostic capabilities of AI/ML models in cancer detection, potentially leading to more accurate and reliable diagnostic tools.

II Methods

The ScienceDirect database was selected due to extensive collection of peer-reviewed articles in AI (Artificial Intelligence) and medical research, ensuring access to high-quality, credible sources.

I searched using keywords relevant to the research focus:

- **"multi-modal"**: To capture studies combining diverse data types (e.g., imaging, genomics);
- **"multi-source data fusion integration"**: To identify frameworks merging heterogeneous datasets;
- **"cancer detection"**: To focus on diagnostic applications;
- **"accuracy"**: To prioritize studies reporting quantitative performance metrics;
- **"multi-modality"**: To ensure coverage of cross-domain integration approaches.

The following search strings were applied:

- `multi-modal AND cancer AND Multi-modality` (777 results);
- `multi-modal AND data fusion integration AND cancer detection AND Multi-modality` (124 results);
- `multi-modal AND multi-source data fusion integration AND cancer detection AND accuracy AND Multi-modality` (89 results).

The full-text search option was enabled to all search strings, and filters limited results to English-language open-access articles published in 2025 to prioritize the most recent advancements, given the rapid evolution of AI technologies.

A reference manager (Zotero) was used to organize citations and ensure consistent formatting.

From the initial 89 articles, I conducted a two-stage screening process. First, I screened titles and abstracts to exclude studies unrelated to AI-driven cancer detection or

multi-modal integration, resulting in 21 articles for further evaluation. Then, in the full-text screening stage, I prioritized two key criteria. Credibility was assessed based on journal CiteScore, peer review, and authors' h-index. Relevance was determined by whether studies used multi-modal data for cancer detection and reported concrete accuracy metrics. I excluded articles lacking clear conclusions about model accuracy or those relying on single-data sources to ensure that only studies directly addressing the research question were included.

Key findings from the ten selected studies were systematically documented in a reading log, facilitating structured analysis, and comparison.

This process received ten articles for final inclusion.

III Results

According to the literature, multi-modal integration significantly enhances diagnostic accuracy across various cancer types by incorporating diverse data sources, including genomic, imaging, and clinical data [4], [5], [6].

While GeneXAI focused on genomic-driven classification [4], OmniFuse [5] and CAFNet [6] prioritized imaging and adaptability to imperfect data, yet all demonstrated superior performance over single-modality approaches, with intermediate fusion strategies like CAFNet [6] offering a balanced synthesis of multimodal inputs.

As for impact on specific cancer types, the evidence showed that multi-modal integration significantly improves diagnostic outcomes across lung, breast, skin, oral, pancreatic, and osteosarcoma cancers by leveraging diverse data sources such as imaging, genomic, and clinical records [6], [7], [8], [9], [10].

Lung and breast cancer studies [7], [8] emphasized broad diagnostic precision through imaging and genomics, whereas oral cancer [6] utilizes fluorescence imaging, a computational pathology framework [9] predicts molecular subtypes using multi-scale histomorphology features from whole slide images, and pancreatic cancer [10] employs protein structures.

The literature underscored that multi-modal integration and tailored dimensionality reduction boosted diagnostic accuracy across cancers by leveraging diverse data and optimizing feature spaces [11], [12], [13]. Dehbozorgi et al. [11] showed pre-trained CNNs (e.g., ResNet50) achieving 96% sensitivity in chest X-ray classification (e.g., pneumonia), emphasizing deep learning's efficiency in multimodal imaging. Albuquerque et al. [12] highlighted object detection models like Faster R-CNN excelling in lesion localization, surpassing human performance with multi-modal imaging. Gliozzo et al. [13] presented a tailored DR pipeline for multi-omics data (e.g., TCGA datasets), enhancing fusion algorithms and survival prediction accuracy using interpretable classifiers.

TABLE I
Risk of Bias Assessment.

Article	**Bias Type**
[4]	Data Source Selection
[5]	Hyperparameter Tuning
[6]	Data Source Selection
[7]	Performance
[8]	Performance Metrics
[9]	Validation Bias
[10]	Small Sample Size
[11]	Generalization Bias
[12]	Reporting Bias
[13]	Class Imbalance in Datasets

After risk of bias analysis two articles [5], [7] with insignificant risk of bias assessment were found. Another eight articles had a significant risk of bias assessment.

IV Discussion

[4], [5], [6], [7], [8], [9], [10], [11], [12], [13] consistently demonstrate that integrating multi-modal data - such as imaging, genomic, proteomic, and clinical information - enhances the accuracy of artificial intelligence and machine learning (AI/ML) models in detecting various cancers, including lung, pancreatic, breast, skin, and osteosarcoma. Relative to single-modality approaches, multi-modal integration resulted in accuracy improvements ranging from 10% to 15.5%. Specific examples include [4], which achieved an accuracy of 95.61% for breast cancer detection using clinical and genomic data, and [5], which reported an area under the receiver operating characteristic curve (AUROC) of 0.894 for lung cancer by combining positron emission tomography (PET), computed tomography (CT), and clinical reports. Similarly, [10] attained an accuracy of 87.5% for pancreatic cancer through the integration of protein and clinical data. These findings highlight the potential of multi-modal strategies to improve diagnostic precision across diverse cancer types.

The results from [4], [5], [6], [7], [8], [9], [10], [11], [12], [13] align with findings from Di Jin et al. [14], who explored multi-modal data integration for pancreatic cancer detection. For instance, according to [10], using of protein and clinical data helped to achieve 87.5% accuracy in pancreatic cancer detection, while [4] reported 95.61% accuracy for breast cancer with clinical and genomic data, and study [5] reached an AUROC of 0.894 for lung cancer using PET, CT, and clinical reports, which supports the approach proposed by Di Jin et al. [14] showing that multimodal methods improve diagnostic accuracy in different cancer types.

The effectiveness of multi-modal integration stems from the complementary strengths of disparate data types. Imaging modalities provide detailed anatomical and structural insights, genomic data elucidates molecular and genetic characteristics, and clinical records contribute patient-specific contextual information, such as medical history and risk factors. This synergy enables AI/ML models to leverage a more comprehensive feature set, thereby improving predictive accuracy. For instance, [5] demonstrated that combining PET and CT imaging captured both metabolic and structural features of lung tumors, enhancing diagnostic reliability. This advantage contrasts sharply with single-modality approaches, which, as

noted in the introduction, often suffer from limited feature representation that hinders performance in complex diagnostic scenarios [1]. However, the observed variation in accuracy gains - from 10% to 15.5% - suggests that the benefits of multi-modal integration are not uniform. Potential contributing factors include differences in data quality, the specific integration techniques employed (e.g., early fusion versus intermediate fusion), and the inherent complexity of the cancer types under investigation. Cancers with greater heterogeneity, such as pancreatic cancer, may derive more pronounced benefits from multi-modal approaches due to the necessity of diverse data inputs to capture their multifaceted nature.

The risk-of-bias assessment highlighted limitations of [4], [5], [6], [7], [8], [9], [10], [11], [12], [13]. The significant limitations are presented below:

In [4], data source selection limited generalizability due to focusing on a single cancer subtype. Future works should include diverse datasets.

In [6], small sample size and limited diversity reduce reliability. Future works should expand datasets with confirmed demographics.

In [8], imbalanced metrics bias evaluations. Future works should report F1/AUC-ROC and confidence intervals.

In [9], validation bias persists due to incompetent training/test splits (e.g., 70%/30%) and lack of independent groups for datasets like OV-EMT and COLU-KRAS. Future works should validate on separate datasets.

In [10], high-risk overfitting from 16-patient data. Future works should prioritize larger datasets.

In [11], hardware reliance limits reproducibility. Future works should optimize for low-resource settings.

In [12], inconsistent metrics make comparison difficult. Future works should adopt standardized reporting.

In [13], class imbalance skews results. Future works should address imbalance via resampling or weighted metrics.

The search strategy was confined to articles published in 2025, a decision that ensures recency but risks excluding foundational studies from earlier years that may have shaped current multi-modal integration techniques. This temporal restriction could result in an incomplete portrayal of the field's development and potentially overlook critical methodological advancements. Additionally, the inclusion of study [8] via snowballing rather than a systematic search introduces the possibility of selection bias. Snowballing depends on the reference lists of already included studies, which may favor certain perspectives or methodologies while neglecting others, thereby skewing the review's findings. Moreover, the screening process reduced an initial pool of 89 articles to 10 based on stringent eligibility criteria. While this focused the analysis, it may have excluded studies that, despite not meeting the precise criteria, offered valuable insights into multi-modal integration.

The results emphasize that the integration of multi-modal information greatly enhances the precision of cancer diagnosis, echoing a more comprehensive strategy than single-modality approaches. Practically, this implies that clinicians may integrate imaging, genomics, and clinical information to improve early diagnosis and personalize treatment plans. Future research needs to prioritize bias overcoming, dataset enlargement, and model optimization to broaden applicability, so that improvements are converted into actual clinical gains. Blending different types of information allows doctors to more readily spot cancer, but there remains much work to be done in making such technology accessible and accurate to everyone.

References

[1] B. Ozdemir and I. Pacal, "An innovative deep learning framework for skin cancer detection employing ConvNeXtV2 and focal self-attention mechanisms," *Results in Engineering*, vol. 25, p. 103692, Mar. 2025. Accessed date: Mar. 9, 2025. DOI. [Online]. Available: https://doi.org/10.1016/j.rineng.2024.103692

[2] A. Uwimana, G. Gnecco, and M. Riccaboni, "Artificial intelligence for breast cancer detection and its health technology assessment: A scoping review," *Computers in Biology and Medicine*, vol. 184, p. 109391, Nov. 2024. Accessed date: Mar. 9, 2025. DOI. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2024.109391

[3] A. Gyasi-Agyei, "A comparative assessment of machine learning models and algorithms for osteosarcoma cancer detection and classification," *Healthcare Analytics*, vol. 7, p. 100380, Jun. 2025. Accessed date: Mar. 9, 2025. DOI. [Online]. Available: https://doi.org/10.1016/j.health.2024.100380

[4] S. Manna, S. Mistry, and D. De, "GeneXAI: Influential Gene Identification For Breast Cancer Stages Using XAI-Based Multi-Modal Framework," *Medicine in Novel Technology and Devices*, pp. 100349–100349, Dec. 2024. Accessed date: Mar. 9, 2025. DOI. [Online]. Available: https://doi.org/10.1016/j.medntd.2024.100349

[5] Y. Wu, J. Chen, L. Hu, H. Xu, H. Liang, and J. Wu, "OmniFuse: A general modality fusion framework for multi-modality learning on low-quality medical data," *Information Fusion*, pp. 102890–102890, Dec. 2024. Accessed date: Mar. 9, 2025. DOI. [Online]. Available: https://doi.org/10.1016/j.inffus.2024.102890

[6] W. Lian, J. Lindblad, C. R. Stark, J.-M. Hirsch, and N. Sladoje, "Let it shine: Autofluorescence of Papanicolaou-stain improves AI-based cytological oral cancer detection," *Computers in Biology and Medicine*, vol. 185, pp. 109498–109498, Dec. 2024. Accessed date: Mar. 9, 2025. DOI. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2024.109498

[7] F. Krones, U. Marikkar, G. Parsons, A. Szmul, and A. Mahdi, "Review of multimodal machine learning approaches in healthcare," *Information Fusion*, pp. 102690–102690, Sep. 2024. Accessed date: Mar. 9, 2025. DOI. [Online]. Available: https://doi.org/10.1016/j.inffus.2024.102690

[8] L. Schneider et al., "Integration of deep learning-based image analysis and genomic data in cancer pathology: A systematic review," *European Journal of Cancer*, vol. 160, pp. 80–91, Jan. 2022. Accessed date: Mar. 9, 2025. DOI. [Online]. Available: https://doi.org/10.1016/j.ejca.2021.10.007

[9] Y. Hu et al., "Self-interactive learning: Fusion and evolution of multi-scale histomorphology features for molecular traits prediction in computational pathology," *Medical Image Analysis*, vol. 101, pp. 103437–103437, Jan. 2025. Accessed date: Mar. 9, 2025. DOI. [Online]. Available: https://doi.org/10.1016/j.media.2024.103437

[10] B. KalaiSelvi, P. Anandan, S. V. Easwaramoorthy, and J. Cho, "IoT-driven cancer prediction: Leveraging AI for early detection of protein structure variations," *Alexandria Engineering Journal*, vol. 118, pp. 21–35, Apr. 2025. Accessed date: Mar. 9, 2025. DOI. [Online]. Available: https://doi.org/10.1016/j.aej.2025.01.038

[11] P. Dehbozorgi, O. Ryabchykov, and T. W. Bocklitz, "A comparative study of statistical, radiomics, and deep learning feature extraction techniques for medical image classification in optical and radiological modalities," *Computers in Biology and Medicine*, vol. 187, pp. 109768–109768, Feb. 2025. Accessed date: Mar. 9, 2025. DOI. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2025.109768

[12] C. Albuquerque, R. Henriques, and M. Castelli, "Deep learning-based object detection algorithms in medical imaging: Systematic review," *Heliyon*, vol. 11, no. 1, p. e41137, Jan. 2025. Accessed date: Mar. 9, 2025. DOI. [Online]. Available: https://doi.org/10.1016/j.heliyon.2024.e41137

[13] J. Gliozzo et al., "Intrinsic-dimension analysis for guiding dimensionality reduction and data fusion in multi-omics data processing," *Artificial Intelligence in Medicine*, vol. 160, p. 103049, Feb. 2025. Accessed date: Mar. 9, 2025. DOI. [Online]. Available: https://doi.org/10.1016/j.artmed.2024.103049

[14] D. Jin, N. U. Khan, W. Gu, H. Lei, A. Goel, and T. Chen, "Informatics strategies for early detection and risk mitigation in pancreatic cancer patients," *Neoplasia*, vol. 60, pp. 101129–101129, Jan. 2025. Accessed date: Mar. 9, 2025. DOI. [Online]. Available: https://doi.org/10.1016/j.neo.2025.101129