RESEARCH PAPER

Graph Intelligence in Oncology: A Comprehensive Review of Patient Similarity Network-Based Cancer Diagnostics

Rajalakshmi Sankaran[1*], Gurumoorthy G[2], P. Padmapriya[3], E. Ramya[4]
[1]Department of Computer Science and Engineering, Sri Venkateswara College of Engineering, Pennalur, Tamil Nadu, India.
[2]Department of Medical Electronics, Saveetha Engineering College, Chennai, Tamil Nadu, India.
[3]Department of Biomedical Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[4]Department of Artificial Intelligence and Data Science, Karpaga Vinayaga Educational Lore for Learning affiliation along with Karpaga Vinayaga College of Engineering and Technology, Chengalpattu, Tamil Nadu, India.
***Corresponding Author:** Rajalakshmi Sankaran, Associate Professor, Department of Computer Science and Engineering, Sri Venkateswara College of Engineering, Pennalur, Tamil Nadu, India. | Email: sraji@svce.ac.in

Abstract

Accurate cancer diagnosis remains a major challenge due to the heterogeneity and high dimensionality of clinical, genomic, and imaging data. Conventional machine learning approaches often fail to capture complex inter-patient relationships, limiting their predictive capability. Graph Neural Networks (GNNs) have recently emerged as an effective solution by modeling data in the form of patient similarity networks, where patients are represented as nodes and their relationships are defined through similarity measures. This review presents a comprehensive analysis of GNN-based cancer diagnostic frameworks that leverage patient similarity networks for improved representation learning. It examines key architectures, including Graph Convolutional Networks and Graph Attention Networks, and discusses methods for constructing similarity graphs using clinical and multimodal data. The review highlights the advantages of these approaches in enhancing diagnostic accuracy and supporting personalized medicine. In addition, it critically analyzes challenges such as scalability, interpretability, data imbalance, and privacy concerns. Recent advancements in explainable GNNs and multimodal integration are also discussed to identify emerging research directions. This study provides a structured overview of current developments and outlines future opportunities for deploying graph-based intelligence in real-world oncology applications.

Keywords

How to cite this article: Rajalakshmi Sankaran1* Gurumoorthy G2 P. Padmapriya3 E. Ramya4. Graph Intelligence in Oncology: A Comprehensive Review of Patient Similarity Network-Based Cancer Diagnostics. Int J Drug Deliv Technol. 2026;16(89s):2129-2139. DOI: 10.25258/ijddt.16.89s.201.

1. Introduction

Cancer remains one of the leading causes of mortality worldwide, necessitating accurate and early diagnostic techniques to improve survival rates and treatment outcomes. The rapid growth of biomedical data, including electronic health records, genomic sequencing, and medical imaging, has enabled the adoption of data-driven approaches in oncology. However, the heterogeneous and high-dimensional nature of these datasets poses significant challenges for conventional machine learning methods, which often fail to capture complex relationships among patients and disease characteristics [1].

1.1 Challenges in Conventional Cancer Diagnosis Approaches

Traditional cancer diagnosis methods rely on statistical models and machine learning techniques that treat patient samples independently. While these approaches have shown promise, they often ignore the inherent relationships among patients, such as shared genetic mutations or similar clinical profiles. This limitation reduces their ability to generalize across diverse populations. Furthermore, challenges such as data imbalance, missing values, and noise significantly impact model performance, particularly in real-world clinical datasets [2].

1.2 Emergence of Graph-Based Learning in Healthcare

Graph-based learning has emerged as an effective paradigm for representing complex and interconnected biomedical data. In graph structures, nodes represent entities such as patients, while edges capture relationships or similarities between them. This representation enables the modeling of dependencies that are otherwise overlooked in traditional approaches. In healthcare, graph-based models have been successfully applied to disease prediction, drug discovery, and patient stratification, demonstrating their ability to capture meaningful patterns in complex datasets [3].

1.3 Graph Neural Networks for Cancer Diagnosis

Graph Neural Networks (GNNs) extend deep learning techniques to graph-structured data, allowing for efficient representation learning by aggregating information from neighboring nodes. Architectures such as Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) have shown significant improvements in cancer diagnosis tasks, including tumor classification and survival prediction. These models leverage both node features and graph topology, resulting in enhanced predictive performance compared to traditional machine learning methods [4].

1.4 Role of Patient Similarity Networks

Patient similarity networks are central to graph-based cancer diagnosis frameworks. In these networks, patients are represented as nodes, and edges are established based on similarity metrics derived from clinical, genomic, or imaging features. This approach enables the grouping of patients with similar characteristics, facilitating more accurate and context-aware predictions. By incorporating relational information, patient similarity networks enhance the capability of GNNs to identify disease patterns and support personalized treatment strategies [5].

1.5 Integration of Multimodal Biomedical Data

The integration of multimodal data is crucial for capturing the complexity of cancer. Combining clinical data with genomic and imaging information allows for a more comprehensive understanding of disease mechanisms. Graph-based models provide a flexible framework for multimodal data fusion, enabling the incorporation of diverse data types into a unified representation. This integration improves diagnostic accuracy and supports the development of precision oncology solutions tailored to individual patients [6].

1.6 Scope and Contributions of the Review

This review presents a comprehensive analysis of Graph Neural Network-based cancer diagnosis using patient similarity networks. It examines recent advancements in graph-based learning techniques, highlighting their strengths and limitations in clinical applications. The study also identifies key challenges, including scalability, interpretability, and data privacy concerns. Furthermore, it explores emerging research directions such as explainable AI and federated learning for healthcare applications. By synthesizing current developments, this review aims to provide valuable insights for researchers and practitioners working toward the development of robust and clinically applicable diagnostic systems [7].

2. Background and Fundamentals

2.1 Graph Neural Networks: Concepts and Architecture

Graph Neural Networks (GNNs) have emerged as a powerful class of deep learning models specifically designed to operate on graph-structured data, where relationships between entities are as important as the entities themselves. Unlike traditional machine learning models that assume independent and identically distributed (i.i.d.) data, GNNs explicitly model dependencies among nodes through a process known as message passing or neighborhood aggregation. In this process, each node updates its feature representation by aggregating information from its neighboring nodes across multiple layers, enabling the capture of both local and global structural patterns [8]–[10].

Figure 1 illustrates the general architecture of Graph Neural Networks (GNNs), highlighting the key stages involved in graph-based learning. The process begins with an input graph G=(V,E,X)G = (V, E, X)G=(V,E,X), where VVV represents the set of nodes (e.g., patients), EEE denotes the edges capturing relationships (e.g., similarity), and XXX corresponds to node feature representations.

The initial node features are passed through multiple GNN layers, where **message passing** is performed. In each layer, nodes aggregate information from their neighboring nodes using aggregation functions such as mean, sum, or attention-based weighting. This aggregated information is then combined with the node's existing features through an update function, enabling the learning of more informative representations. The diagram further illustrates different GNN variants, including Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), and GraphSAGE. GCNs perform convolution operations on graph structures, while GATs incorporate attention mechanisms to assign varying importance to neighboring nodes. GraphSAGE introduces neighborhood sampling techniques to improve scalability for large graphs.

After multiple iterations, the model generates **node embeddings**, which capture both structural and feature-based information. These embeddings are then used in the prediction layer for downstream tasks such as node classification, graph classification, and link prediction. In the context of cancer diagnosis, these outputs correspond to tasks such as tumor classification, subtype identification, and treatment response prediction.

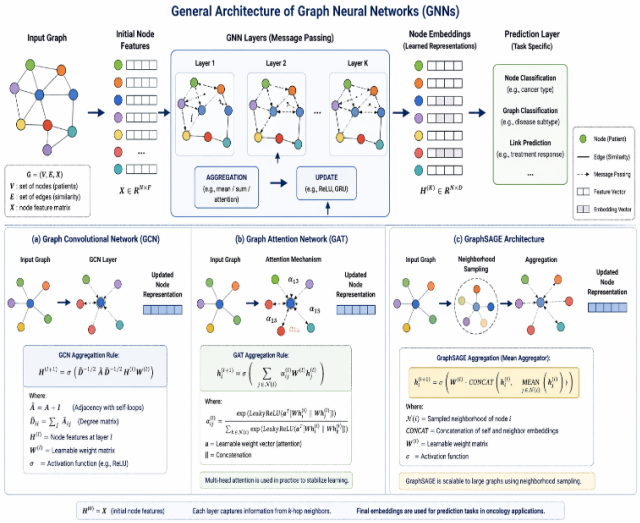


Figure 1 *General architecture of Graph Neural Networks*

In oncology applications, the nodes typically represent patients, while edges encode similarity relationships derived from clinical, genomic, or imaging data, forming the basis of patient similarity networks used for diagnosis and prognosis. The fundamental architecture of a GNN consists of three key components: (i) **node feature initialization**, (ii) **aggregation function**, and (iii) **update function**. During each iteration, node features are combined with those of their neighbors using aggregation strategies such as mean, sum, or attention-weighted mechanisms. This iterative propagation allows the network to learn high-level embeddings that encode both node attributes and graph topology.

Among the widely used GNN variants, **Graph Convolutional Networks (GCNs)** perform spectral or spatial convolutions on graphs, making them efficient for semi-supervised learning tasks. **Graph Attention Networks (GATs)** introduce attention mechanisms to assign varying importance to neighboring nodes, thereby improving the model's ability to focus on relevant relationships. **GraphSAGE** extends GNNs to large-scale graphs by sampling neighborhoods instead of processing the entire graph, enhancing scalability.

In the context of oncology, GNNs are particularly valuable because biomedical systems inherently exhibit relational structures, such as gene interaction networks, protein-protein interaction networks, and patient similarity graphs. By leveraging these relationships, GNNs can uncover hidden patterns that are not accessible through traditional models, thereby improving diagnostic accuracy and robustness.

2.2 Patient Similarity Networks

Patient Similarity Networks (PSNs) provide a graph-based framework for modeling relationships among patients based on shared clinical, genomic, or phenotypic characteristics. In a PSN, each node represents an individual patient, while edges represent the degree of similarity between patients, quantified using predefined or learned similarity metrics. These metrics may include Euclidean distance, cosine similarity, correlation coefficients, or more advanced embedding-based measures derived from deep learning models [11], [12]. The construction of PSNs typically involves three steps: (i) feature extraction, where relevant patient attributes are collected from multiple data sources; (ii) similarity computation, where pairwise relationships are quantified; and (iii) graph formation, where edges are established based on similarity thresholds or k-nearest neighbor strategies. The resulting network captures both

explicit and implicit relationships among patients, enabling more effective learning.

In cancer diagnosis, PSNs play a crucial role in addressing disease heterogeneity. Cancer is not a single disease but a collection of disorders with diverse molecular and clinical characteristics. By grouping patients with similar profiles, PSNs facilitate patient stratification, subtype discovery, and personalized treatment planning. Moreover, integrating PSNs with GNNs allows models to propagate information across similar patients, improving predictions for cases with limited or noisy data.

Another important advantage of PSNs is their ability to incorporate multimodal similarity, where relationships are defined based on a combination of clinical, genomic, and imaging features. This leads to a more comprehensive representation of patient characteristics and enhances the model's ability to capture complex disease patterns.

2.3 Data Sources in Cancer Diagnosis

Cancer diagnosis relies on a wide range of heterogeneous biomedical data sources, each contributing unique insights into disease mechanisms and progression. The primary data modalities include:

- **Clinical Data:** Electronic Health Records (EHRs), patient demographics, laboratory test results, and treatment histories. These data provide contextual and longitudinal information about patient health.
- **Genomic Data:** DNA sequencing, gene expression profiles, and mutation data. Genomic information plays a critical role in understanding the molecular basis of cancer and identifying biomarkers.
- **Medical Imaging Data:** Radiological images such as MRI, CT scans, and histopathological slides. Imaging data enable the visualization of tumor morphology and spatial characteristics.
- **Proteomic and Multi-Omics Data:** Protein expression, metabolomics, and epigenetic data, which offer deeper insights into cellular processes and disease pathways.

The integration of these diverse data sources is essential for accurate cancer diagnosis but presents significant challenges due to differences in data formats, scales, and noise levels. Traditional models often struggle to effectively combine such heterogeneous data. However, graph-based approaches, particularly GNNs, provide a flexible and scalable framework for multimodal data fusion, where different data types can be represented as nodes, edges, or attributes within a unified graph structure [13], [14].

Furthermore, the use of multimodal data enhances the robustness and generalizability of diagnostic models by capturing complementary information across different sources. For example, combining genomic and imaging data can improve tumor classification accuracy, while integrating clinical and molecular data supports better prognosis prediction. Despite these advantages, challenges such as data sparsity, missing values, and privacy concerns must be addressed to ensure reliable model performance and clinical applicability. Ongoing research focuses on developing advanced data integration techniques and privacy-preserving frameworks to overcome these limitations.

2.4 Deep Learning-Based Cancer Diagnosis and Medical Image Analysis

Sinthia et al. (2026) proposed a synergistic deep learning fusion approach for precision lung cancer staging using computed tomography (CT) images. The study employed image preprocessing, lung segmentation, data augmentation, and a customized convolutional neural network to extract discriminative imaging features and classify lung cancer into stages I–IV. The reported results demonstrated promising classification performance, with an overall accuracy of 93.0% and strong area-under-the-curve values across the four stages. This study highlights the potential of deep learning-based feature extraction to support automated cancer staging. Integrating such imaging-derived features with patient similarity networks could further support the identification of clinically similar patients and enhance graph-based diagnostic analysis.

Mohan et al. (2026) introduced an Inception U-Net-based deep learning method for brain tumour segmentation, combining Inception modules with feature extraction techniques to improve the identification of tumour regions in brain images. The study evaluated segmentation performance using the BraTS 2018 datasets and reported statistically significant improvements across glioblastoma subregions. Accurate tumour segmentation can provide valuable information about tumour location, morphology, and extent. These features could be incorporated into patient representations within a patient similarity network, enabling the exploration of relationships among patients with comparable tumour characteristics and supporting graph-based clinical decision-making.

Although deep learning-based imaging methods provide valuable diagnostic features, patient similarity networks offer a complementary framework for representing relationships among patients based on their clinical, imaging, and molecular characteristics.

3. GNN-Based Cancer Diagnosis Approaches

3.1 Graph Construction Techniques

Graph construction is a foundational step in GNN-based cancer diagnosis, as it directly determines how relationships between patients or biological entities are represented and learned. The quality and structure of the graph significantly influence the effectiveness of downstream learning tasks such as classification, clustering, and prediction [15].

Figure 2 illustrates the major graph construction strategies used in GNN-based cancer diagnosis. Feature-based similarity graphs connect patients based on shared attributes, while distance-based graphs such as k-nearest neighbor networks ensure sparsity by linking only the most similar nodes. Biological knowledge-based graphs incorporate domain-specific relationships such as gene interactions and protein networks. These techniques highlight the importance of graph design in capturing meaningful relationships and improving model performance.

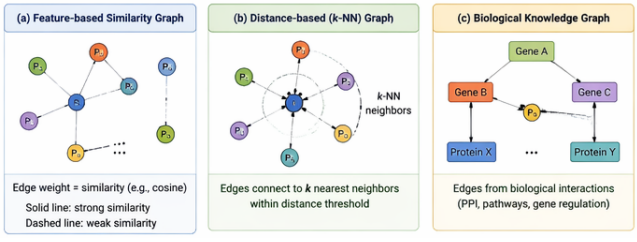


Figure 2 *Different graph construction techniques for patient similarity networks.*

One of the most commonly used approaches is **feature-based similarity graph construction**, where edges between nodes are defined based on similarity measures computed from patient features. These features may include clinical attributes (age, symptoms, lab results), genomic profiles (gene expression, mutations), or imaging-derived features. Similarity metrics such as cosine similarity, Euclidean distance, and Pearson correlation are typically used to quantify relationships. A threshold is then applied to

determine whether an edge should exist between two nodes. While this approach is intuitive and easy to implement, it is sensitive to noise and feature scaling.

Another widely adopted technique is **distance-based graph construction**, particularly the k-nearest neighbor (k-NN) method. In this approach, each node is connected to its k most similar neighbors based on a chosen distance metric. This ensures a sparse graph structure, which reduces computational complexity and prevents overfitting. However, selecting an appropriate value of k is critical, as too small a value may lead to information loss, while too large a value may introduce noisy connections. In contrast, **biological knowledge-based graph construction** incorporates domain-specific information such as gene regulatory networks, protein-protein interaction (PPI) networks, and metabolic pathways. These graphs leverage prior biological knowledge to establish meaningful relationships between nodes, thereby enhancing model interpretability and relevance. Such approaches are particularly beneficial in cancer research, where biological interactions play a crucial role in disease progression.

More advanced methods involve **heterogeneous graph construction**, where multiple types of nodes (patients, genes, proteins) and edges (similarity, interaction, regulation) are integrated into a single graph. This allows for a richer representation of complex biomedical systems and supports more comprehensive learning. However, it also increases model complexity and computational requirements.

3.2 GNN Architectures in Oncology

The choice of GNN architecture plays a critical role in determining the performance and scalability of cancer diagnostic models. Different architectures are designed to address specific challenges associated with graph learning, such as handling large-scale data, capturing complex relationships, and improving interpretability [8], [16].

Figure 3 presents a comparative overview of key GNN architectures, including GCN, GAT, and GraphSAGE. Each architecture differs in its aggregation and learning mechanisms. GCNs perform uniform aggregation, GATs introduce attention-based weighting, and GraphSAGE focuses on scalability through neighborhood sampling. These differences influence their suitability for various cancer diagnosis tasks, such as classification, prediction, and clustering.

Architecture	Key Idea	Mathematical Operation (One Layer)	Strengths	Applications in Oncology
(a) GCN	Convolution in graph spectral domain	$H^{(l+1)} = \sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right)$ $\hat{A} = A + I$ $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$	• Efficient • Simple • Good for semi-supervised tasks	Cancer type classification, mutation prediction, survival analysis
(b) GAT	Attention mechanism to learn importance of neighbors	$h_i^{(l+1)} = \sigma\left(\sum_{j\in\mathcal{N}(i)} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l)}\right)$ $\alpha_{ij}^{(l)} = \frac{\exp(\text{LeakyReLU}(a^T[Wh_i^{(l)} \| Wh_j^{(l)}]))}{\sum_{k\in\mathcal{N}(i)} \exp(\text{LeakyReLU}(a^T[Wh_i^{(l)} \| Wh_k^{(l)}]))}$	• Adaptive weighting • Captures complex relationships	Patient stratification, biomarker discovery, drug response prediction
(c) GraphSAGE	Inductive learning using neighbor sampling and aggregation	$h_i^{(l+1)} = \sigma\left(W^{(l)} \cdot CONCAT\ h_i^{(l)}, AGGREGATE\left(\{h_j^{(l)}, j\in\mathcal{N}(i)\}\right)\right)$	• Scalable • Works on large graphs • Inductive capability	Large-scale patient networks, multi-center datasets, real-world clinical deployment

Figure 3 *Comparison of major Graph Neural Network architectures used in oncology applications.*

Graph Convolutional Networks (GCNs) are among the earliest and most widely used GNN architectures. They perform convolution operations on graph data by aggregating features from neighboring nodes and applying linear transformations. GCNs are particularly effective for semi-supervised learning tasks, where only a subset of nodes is labeled. In oncology, GCNs have been successfully applied to tumor classification and disease prediction tasks.

Graph Attention Networks (GATs) extend GCNs by introducing attention mechanisms that assign different weights to neighboring nodes based on their importance. This enables the model to focus on more relevant relationships, improving performance in heterogeneous datasets. GATs are especially useful in cancer diagnosis, where certain patient similarities or gene interactions may carry more significance than others.

GraphSAGE is designed for scalability and inductive learning. Instead of processing the entire graph, it samples a subset of neighbors for each node during training. This makes it suitable for large-scale healthcare datasets, such as those involving thousands of patients across multiple institutions. In addition to these standard architectures, **hybrid models** combining GNNs with Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs) have been developed. For example, CNNs can be used to extract features from medical images, which are then incorporated into a graph structure for relational learning using GNNs. These hybrid approaches have demonstrated improved performance in complex cancer diagnosis tasks.

3.3 Multimodal GNN Models

Cancer diagnosis often requires the integration of multiple data modalities, each capturing different aspects of the disease. Multimodal GNN models address this need by combining diverse data sources into a unified graph-based framework, enabling more comprehensive and accurate predictions [17].

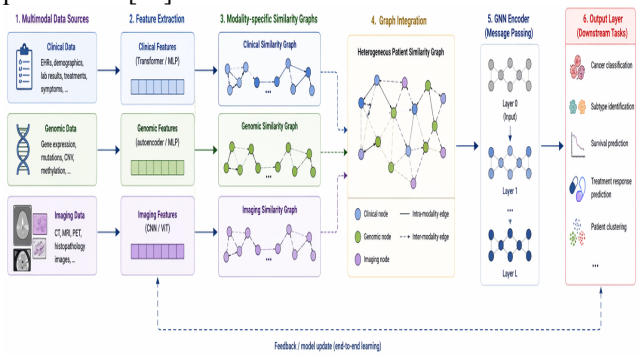


Figure 4 *Multimodal GNN framework for cancer diagnosis using patient similarity networks.*

Figure 4 illustrates the workflow of a multimodal GNN model, where clinical, genomic, and imaging data are first processed to extract features. Separate similarity graphs are constructed for each modality and then integrated into a unified graph. The combined graph is passed through multiple GNN layers to learn node embeddings, which are used for downstream tasks such as cancer classification, subtype identification, and survival prediction. This framework demonstrates how multimodal integration enhances diagnostic performance.

One approach to multimodal integration is **early fusion**, where features from different modalities are combined into a single feature vector before graph construction. While this approach is simple, it may fail to capture modality-specific patterns effectively.

Alternatively, **late fusion** involves constructing separate graphs for each modality (e.g., clinical, genomic, imaging) and combining their outputs at the representation level. This allows each modality to be processed independently while still contributing to the final prediction. A more advanced strategy is **hybrid or heterogeneous graph construction**, where nodes represent different entities (patients, genes, images) and edges capture relationships both within and

across modalities. This enables the model to learn complex interactions between different data types.

Recent advancements include the use of **cross-modal attention mechanisms**, which allow the model to dynamically assign importance to different modalities based on the task. This improves performance by focusing on the most relevant information.

Multimodal GNNs have shown significant improvements in cancer subtype classification, prognosis prediction, and treatment response analysis. By leveraging complementary information from multiple sources, these models provide a more holistic understanding of disease patterns and support precision oncology.

4. Applications in Cancer Diagnosis

Graph Neural Networks (GNNs), when combined with Patient Similarity Networks (PSNs), have demonstrated significant advancements across multiple cancer diagnosis tasks. Their ability to model relationships among patients and integrate heterogeneous biomedical data makes them particularly effective in addressing the complexity and variability of cancer. This section explores the major application areas in detail.

4.1 Cancer Type Classification

Cancer type classification is one of the primary applications of GNN-based models, where the objective is to accurately categorize tumors into predefined classes based on patient data. Traditional machine learning models typically treat each sample independently, relying solely on feature vectors without considering relationships between patients. This limitation often leads to reduced performance, especially in cases where patient data is sparse or noisy.

GNN-based approaches overcome this limitation by leveraging relational dependencies among patients through graph structures. In a patient similarity network, patients with similar clinical or molecular profiles are connected, allowing information to propagate across nodes. This enables the model to learn contextual representations that capture both individual features and neighborhood information. For example, in cancer type classification tasks, GNNs can integrate genomic mutations, gene expression profiles, and clinical attributes into a unified representation. By aggregating information from similar patients, the model can improve classification accuracy, particularly for rare cancer types where labeled data is limited. Additionally, GNNs are capable of handling class imbalance more effectively by leveraging neighborhood information to compensate for underrepresented classes.

Recent studies have demonstrated that GNN-based classifiers outperform conventional approaches such as Support Vector Machines (SVMs) and Random Forests in terms of accuracy, robustness, and generalization ability [18]. Furthermore, the incorporation of attention mechanisms allows the model to identify the most influential features and relationships, enhancing interpretability.

4.2 Cancer Subtype Identification

Cancer subtype identification is a critical task in precision oncology, as different subtypes of the same cancer often exhibit distinct biological characteristics and treatment responses. Traditional clustering methods, such as k-means or hierarchical clustering, rely primarily on feature similarity and often fail to capture complex interactions among patients. Patient Similarity Networks (PSNs) combined with GNNs provide a more effective approach by modeling both feature-based and relational similarities. In this framework, patients are grouped into clusters based on their connectivity within the graph, which reflects shared molecular and clinical characteristics.

GNNs enhance subtype identification by learning embeddings that preserve both local neighborhood structure and global graph topology. These embeddings can then be used for clustering, enabling the discovery of previously unknown cancer subtypes. For instance, in breast cancer and lung cancer studies, GNN-based models have successfully identified subgroups with distinct genomic signatures and survival outcomes. Moreover, PSN-based clustering supports **patient stratification**, which is essential for personalized treatment planning. By identifying subtypes associated with specific biomarkers or genetic mutations, clinicians can tailor therapies to individual patients, improving treatment efficacy and reducing adverse effects [11].

Another advantage of GNN-based subtype identification is its ability to incorporate multimodal data, allowing for a more comprehensive understanding of disease heterogeneity. This is particularly important in oncology, where cancer progression is influenced by multiple biological factors.

4.3 Prognosis and Survival Prediction

Prognosis and survival prediction involve estimating the likelihood of disease progression, recurrence, or patient survival over time. These tasks are inherently complex, as they depend on multiple factors, including tumor characteristics, treatment history, and patient-specific conditions. Traditional survival analysis methods, such as Cox proportional hazards models, assume independence between samples and often fail to capture interactions among patients. GNN-based models address this limitation by incorporating patient relationships into the prediction process.

In a graph-based framework, patients with similar clinical or molecular profiles are connected, enabling the model to propagate survival-related information across the network. This allows the model to learn more robust representations, improving prediction accuracy. For example, patients with similar gene expression patterns and treatment responses may exhibit comparable survival outcomes, which can be captured effectively using GNNs.

Recent approaches integrate survival analysis with GNN architectures by combining node embeddings with hazard prediction models. These models have demonstrated improved performance in predicting overall survival, progression-free survival, and treatment response compared to traditional methods [19]. Additionally, GNNs can handle censored data, which is common in medical datasets, by incorporating survival-specific loss functions. The integration of temporal information further enhances prediction accuracy, enabling dynamic prognosis modeling.

4.4 Medical Imaging-Based Diagnosis

Medical imaging plays a crucial role in cancer diagnosis, providing detailed information about tumor structure, size, and spatial characteristics. However, traditional image analysis methods often rely on pixel-level features and may fail to capture higher-level spatial relationships. Graph-based approaches address this limitation by representing images as graphs, where nodes correspond to image regions (e.g., patches or superpixels), and edges represent spatial or functional relationships. GNNs can then be applied to these graphs to learn spatially-aware representations.

In histopathology, for example, tissue images can be divided into regions, and relationships between regions can be modeled to capture tumor microenvironment interactions.

GNNs can identify patterns such as cell distribution, tissue organization, and morphological changes, which are critical for accurate diagnosis. Similarly, in radiology, GNNs can be used to analyze MRI and CT scans by modeling spatial dependencies between different anatomical regions. This improves tumor detection, segmentation, and classification accuracy.

Another emerging approach involves combining Convolutional Neural Networks (CNNs) with GNNs, where CNNs extract local features from images, and GNNs capture global relationships. This hybrid framework has shown significant improvements in cancer detection and classification tasks [20]. Overall, GNN-based imaging analysis provides a more comprehensive understanding of tumor characteristics by integrating spatial and relational information, leading to improved diagnostic performance.

5. Advantages of GNN-Based Patient Similarity Networks

Graph Neural Network (GNN)-based Patient Similarity Networks (PSNs) offer several compelling advantages over traditional machine learning approaches in cancer diagnosis. By explicitly modeling relationships among patients and integrating heterogeneous biomedical data, these frameworks provide a more comprehensive and context-aware understanding of disease patterns. **Figure 5** illustrates the key advantages of Graph Neural Network (GNN)-based Patient Similarity Networks (PSNs) in cancer diagnosis. The central component represents the PSN framework, where patients are modeled as nodes and relationships between them are defined through similarity measures. This structure enables the learning of complex inter-patient relationships using graph neural networks.

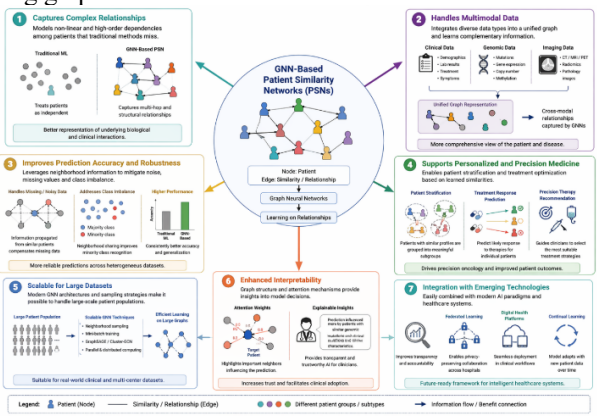


Figure 5 *Advantages of GNN-based patient similarity networks in cancer diagnosis.*

The key advantages are discussed in detail below.

5.1 Capturing Complex Inter-Patient Relationships

One of the most significant advantages of GNN-based PSNs is their ability to capture complex and non-linear relationships among patients. Traditional models typically treat patient samples as independent observations, ignoring the underlying dependencies that may exist due to shared genetic mutations, similar clinical histories, or comparable treatment responses. In contrast, PSNs represent patients as nodes in a graph, with edges encoding similarity relationships. GNNs leverage this structure through iterative message passing, allowing information to propagate across neighboring nodes. This enables the model to learn contextual representations that reflect both individual patient features and their relationships with others.Such relational modeling is particularly important in oncology, where cancer progression is influenced by multiple interacting biological factors. By capturing these interactions, GNN-based models can uncover hidden patterns and associations that are not detectable using conventional approaches, leading to improved diagnostic and predictive performance [9].

5.2 Effective Handling of Multimodal Data

Cancer diagnosis involves diverse data modalities, including clinical records, genomic profiles, and medical imaging. Each modality provides unique and complementary information, but integrating them effectively remains a major challenge for traditional models.

GNN-based PSNs offer a flexible framework for multimodal data integration by representing different data types within a unified graph structure. Features from various modalities can be incorporated as node attributes or used to define edges between nodes. Additionally, heterogeneous graphs allow the inclusion of multiple node and edge types, enabling the modeling of complex interactions across modalities. This capability allows GNNs to capture cross-modal relationships, such as the association between gene mutations and imaging features or the correlation between clinical parameters and treatment outcomes. As a result, multimodal GNN models provide a more holistic understanding of cancer, improving diagnostic accuracy and robustness [16].

5.3 Improved Prediction Accuracy and Robustness

The incorporation of relational information significantly enhances the predictive performance of GNN-based models. By aggregating information from similar patients, GNNs can compensate for missing or noisy data, leading to more reliable predictions. For example, in cases where a patient has incomplete clinical records, the model can leverage information from neighboring patients with similar characteristics to infer missing patterns. This improves the robustness of the model, particularly in real-world clinical settings where data quality is often inconsistent. Moreover, GNNs are effective in handling **class imbalance**, a common issue in cancer datasets where certain cancer types or subtypes are underrepresented. By utilizing neighborhood information, GNNs can improve classification performance for minority classes, reducing bias and enhancing generalization. Empirical studies have consistently shown that GNN-based approaches outperform traditional machine learning models such as decision trees, support vector machines, and shallow neural networks in various cancer diagnosis tasks [9], [16].

5.4 Enabling Personalized and Precision Medicine

Personalized medicine aims to tailor treatment strategies based on individual patient characteristics. GNN-based PSNs play a crucial role in this paradigm by enabling patient stratification and subgroup identification. By clustering patients with similar molecular and clinical profiles, PSNs allow clinicians to identify subgroups that may respond differently to specific treatments. GNNs further enhance this process by learning embeddings that capture both patient features and relational context, enabling more accurate predictions of treatment outcomes. This approach supports **precision oncology**, where therapies are customized based on genetic mutations, tumor characteristics, and patient-specific factors. For instance, GNN models can help identify patients who are likely to benefit from targeted therapies or immunotherapy, improving treatment effectiveness and reducing adverse effects.

5.5 Scalability and Adaptability to Large-Scale Data

With the increasing availability of large-scale biomedical datasets, scalability has become a critical requirement for modern diagnostic models. GNN architectures such as GraphSAGE and sampling-based methods are specifically designed to handle large graphs efficiently.

These models use techniques such as neighborhood sampling and mini-batch training to reduce computational complexity, making them suitable for large patient populations. Additionally, distributed and parallel computing frameworks further enhance scalability, enabling real-time analysis in clinical environments. Another advantage is the adaptability of GNN-based PSNs to evolving datasets. As new patient data becomes available, the graph can be updated dynamically, allowing the model to continuously learn and improve over time. This is particularly important in healthcare, where data is constantly being generated.

5.6 Enhanced Interpretability Through Graph Structures

Although GNNs are often considered complex models, the graph structure itself provides a degree of interpretability. The relationships between patients can be visualized and analyzed, offering insights into how predictions are made. For example, attention mechanisms in GATs can highlight which neighboring nodes contribute most to a prediction, helping clinicians understand the factors influencing the model's decision. This is a step toward building trust in AI-based diagnostic systems.

5.7 Integration with Emerging Technologies

GNN-based PSNs can be seamlessly integrated with emerging technologies such as explainable AI, federated learning, and digital health platforms. This integration enhances their applicability in real-world clinical settings by addressing challenges related to transparency, privacy, and interoperability. Overall, GNN-based patient similarity networks provide a powerful framework for cancer diagnosis by combining relational modeling, multimodal data integration, and scalable learning. Their ability to capture complex dependencies and support personalized treatment strategies makes them a promising tool for advancing precision oncology.

6. Challenges and Limitations

Despite the promising capabilities of Graph Neural Networks (GNNs) combined with Patient Similarity Networks (PSNs), several challenges hinder their widespread adoption in clinical oncology. These limitations arise from both technical and practical considerations and must be addressed to enable reliable real-world deployment.

6.1 Data Quality Issues

One of the most critical challenges in GNN-based cancer diagnosis is the quality and reliability of biomedical data. Clinical datasets often contain missing values due to incomplete patient records, inconsistencies in data collection, and variations in clinical practices across institutions. Similarly, genomic and imaging datasets may include noise, measurement errors, or batch effects that distort underlying biological signals. GNNs rely heavily on neighborhood aggregation, meaning that errors in one node can propagate through the network, potentially amplifying noise and leading to inaccurate predictions. This issue is particularly problematic in patient similarity networks, where incorrect similarity relationships may result in misleading connections between patients.

Moreover, class imbalance is a common issue in cancer datasets, especially for rare cancer types or subtypes. Traditional learning algorithms tend to bias toward majority classes, and although GNNs partially mitigate this through neighborhood information sharing, imbalance still affects performance. To address these issues, robust preprocessing techniques such as data imputation, normalization, and noise filtering are required. Additionally, advanced strategies like robust graph learning and uncertainty-aware GNNs are emerging to improve model resilience [13].

6.2 Scalability

Scalability remains a major limitation for GNN-based models, particularly when dealing with large-scale healthcare datasets involving thousands or millions of patients. Graph construction itself can be computationally expensive, especially when calculating pairwise similarities across high-dimensional feature spaces. Furthermore, the message passing mechanism in GNNs requires repeated aggregation over neighbors, which increases computational complexity with graph size and density. This becomes a bottleneck in real-time clinical applications where fast and efficient processing is essential. Although techniques such as GraphSAGE, sampling-based methods, and mini-batch training have been proposed to address scalability, they often involve trade-offs between computational efficiency and model accuracy. Additionally, memory constraints pose challenges when handling large graphs with rich node features. Future research must focus on developing lightweight GNN architectures, distributed training frameworks, and efficient graph sparsification techniques to enable large-scale deployment in clinical environments [8].

6.3 Interpretability

Interpretability is a crucial requirement in healthcare applications, where decisions must be transparent and explainable to clinicians. However, GNNs are often considered "black-box" models due to their complex architecture and iterative message passing processes. While attention mechanisms in models such as Graph Attention Networks (GATs) provide some level of interpretability, they do not fully explain how predictions are made. Similarly, post-hoc explanation methods like GNNExplainer offer insights into important nodes and edges but may not always align with clinical reasoning. The lack of interpretability limits trust and acceptance among healthcare professionals, as clinicians require clear justifications for diagnostic decisions. Therefore, developing clinically interpretable GNN frameworks that provide meaningful explanations aligned with medical knowledge is essential. Recent advancements in Explainable Artificial Intelligence (XAI) aim to address this issue, but their integration into GNN-based cancer diagnosis systems remains an open research challenge [21].

6.4 Privacy and Ethical Concerns

Cancer diagnosis involves highly sensitive patient data, including medical history, genetic information, and imaging records. The use of such data in machine learning models raises significant **privacy and ethical concerns**, particularly when data is shared across institutions.

Most GNN-based approaches require centralized data processing, which increases the risk of data breaches and unauthorized access. Additionally, regulatory frameworks such as HIPAA and GDPR impose strict requirements on data handling and sharing. Another concern is **bias and fairness**, as models trained on limited or non-representative datasets may produce biased predictions that disproportionately affect certain patient groups.

To address these issues, privacy-preserving techniques such as **federated learning, differential privacy, and secure multi-party computation** are being explored. However, integrating these techniques with GNN frameworks introduces additional complexity and remains an active area of research [22].

7. Emerging Trends and Future Directions

To overcome the existing limitations, several emerging research directions are shaping the future of GNN-based cancer diagnosis.

7.1 Explainable Graph Neural Networks

Explainable GNNs aim to improve the transparency of graph-based models by providing interpretable insights into their decision-making processes. Techniques such as attention visualization, subgraph extraction, and feature attribution help identify the most influential nodes, edges, and features contributing to predictions. In oncology, explainable GNNs can highlight critical biomarkers, patient similarities, or imaging patterns that influence diagnostic outcomes. This not only enhances trust among clinicians but also supports clinical validation of model predictions. Future research is expected to focus on developing domain-specific explanation methods that align with medical knowledge and provide actionable insights for healthcare professionals [21].

7.2 Federated Learning in Healthcare

Federated learning has emerged as a promising solution for addressing privacy concerns by enabling collaborative model training across multiple institutions without sharing raw data. In this framework, models are trained locally on decentralized datasets, and only model parameters are shared and aggregated. Integrating federated learning with GNNs enables the construction of distributed patient similarity networks, where knowledge is shared across institutions while preserving data privacy. This approach is particularly valuable in oncology, where large and diverse datasets are required for robust model training. However, challenges such as communication overhead, model synchronization, and heterogeneity of local datasets must be addressed to ensure efficient implementation [22].

7.3 Multi-Omics and Multimodal Integration

The integration of multi-omics data, including genomics, proteomics, transcriptomics, and metabolomics, represents a significant advancement in cancer diagnosis. GNNs provide a natural framework for modeling complex interactions among these data types through heterogeneous graphs. Future research is likely to focus on cross-modal learning and representation fusion, enabling models to capture interactions between different biological layers. This will lead to a deeper understanding of cancer mechanisms and improved diagnostic accuracy. Additionally, combining multi-omics data with clinical and imaging data will further enhance the predictive capabilities of GNN-based models, supporting precision oncology initiatives [17].

7.4 Real-Time Clinical Deployment

For GNN-based models to be adopted in clinical practice, they must be integrated into real-time decision support systems. This requires addressing challenges related to scalability, latency, and usability.

Emerging solutions include edge computing, cloud-based platforms, and optimized GNN architectures, which enable faster processing and real-time inference. Furthermore, user-friendly interfaces and visualization tools are essential for facilitating interaction between clinicians and AI systems. The development of clinically validated and regulatory-compliant systems will be crucial for bridging the gap between research and practical implementation [16].

8. Conclusion

Graph Neural Networks have introduced a transformative approach to cancer diagnosis by enabling the modeling of patient similarity networks and integrating heterogeneous biomedical data into a unified framework. Their ability to capture complex inter-patient relationships, leverage multimodal data, and improve predictive performance makes them highly effective for advancing precision oncology. However, several challenges, including data quality issues, scalability limitations, lack of interpretability, and privacy concerns, must be addressed to ensure reliable clinical adoption. Ongoing research in explainable AI, federated learning, and multimodal integration is paving the way for more robust and trustworthy GNN-based diagnostic systems. In the future, the convergence of graph intelligence with emerging healthcare technologies is expected to enable real-time, personalized, and data-driven cancer diagnosis. By addressing current limitations and leveraging new advancements, GNN-based patient similarity networks hold significant potential to revolutionize oncology and improve patient outcomes.

9. Research Gap and Problem Statement

9.1 Research Gaps

Despite the rapid advancements in Graph Neural Networks (GNNs) and their successful application in cancer diagnosis, several critical research gaps remain that limit their full potential in real-world clinical settings. Firstly, graph construction remains largely heuristic and domain-dependent. Most existing studies rely on predefined similarity metrics such as Euclidean distance or cosine similarity to construct patient similarity networks. However, these approaches may fail to capture complex, nonlinear relationships among patients, particularly when dealing with high-dimensional and heterogeneous biomedical data. There is a lack of adaptive and data-driven graph construction methods that can dynamically learn optimal relationships.

Secondly, although multimodal data integration has gained attention, existing models often fail to effectively fuse heterogeneous data sources. Many approaches either rely on simple feature concatenation or process modalities independently, which limits the ability to capture cross-modal interactions. The challenge of integrating clinical, genomic, and imaging data into a unified and interpretable framework remains insufficiently addressed. Another significant gap lies in the limited interpretability of GNN models. While GNNs provide high predictive accuracy, they are often considered “black-box” models. In clinical applications, interpretability is crucial for gaining trust from healthcare professionals and ensuring transparency in decision-making. Existing explainable AI techniques for GNNs are still in their early stages and are not fully adapted for complex medical datasets.

Furthermore, scalability and computational efficiency pose major challenges. Many GNN models struggle to handle large-scale patient datasets due to the high computational cost associated with graph construction and message passing. This limits their applicability in real-world healthcare systems where data is continuously growing.

In addition, data quality and imbalance issues significantly affect model performance. Medical datasets often contain missing values, noisy data, and imbalanced class distributions, particularly in rare cancer types. Current GNN-based approaches do not adequately address these challenges, leading to biased or unreliable predictions.

Another important limitation is the lack of standardized benchmarking and evaluation protocols. Different studies use varied datasets, preprocessing techniques, and evaluation metrics, making it difficult to compare model performance objectively. This lack of consistency hinders the reproducibility and generalization of research findings.

Finally, privacy and security concerns remain underexplored. Cancer diagnosis involves sensitive patient

data, and most existing models require centralized data processing, which raises ethical and legal issues. Although federated learning has been proposed as a solution, its integration with GNN frameworks is still limited and requires further investigation.

9.2 Problem Statement

Based on the identified research gaps, there is a clear need for a robust, scalable, and interpretable framework for cancer diagnosis that effectively leverages patient similarity networks and graph-based learning.

The core problem addressed in this study is:

How to design an efficient and interpretable Graph Neural Network-based framework that can dynamically construct patient similarity networks, integrate heterogeneous multimodal biomedical data, and provide accurate and reliable cancer diagnosis while ensuring scalability and data privacy?

To address this problem, future research must focus on developing:

- **Adaptive graph construction methods** that learn similarity relationships directly from data rather than relying on predefined metrics
- **Advanced multimodal fusion techniques** capable of capturing complex interactions between clinical, genomic, and imaging data
- **Explainable GNN models** that provide transparent and clinically interpretable insights
- **Scalable architectures** that can handle large-scale patient datasets efficiently
- **Robust learning strategies** to address data imbalance, noise, and missing values
- **Privacy-preserving frameworks**, such as federated GNNs, for secure healthcare data analysis

Addressing these challenges will enable the development of next-generation cancer diagnostic systems that are not only accurate but also reliable, interpretable, and suitable for real-world clinical deployment.

10. Literature Survey and Comparative Analysis

The comparative analysis presented in **Table 1** highlights the evolving landscape of Graph Neural Network (GNN)-based approaches for cancer diagnosis using patient similarity networks. A clear trend observed across the reviewed studies is the shift from traditional machine learning models toward graph-based frameworks that explicitly capture inter-patient relationships.

Table 1 *Comparative Analysis of GNN-Based Cancer Diagnosis Using Patient Similarity Networks*

S.No	Author & Year	Methodology	Data Type	GNN Model	Key Contribution	Limitations
1	Parisot et al., 2018	PSN + GCN	Clinical + Imaging	GCN	Introduced PSN for disease prediction	Limited scalability
2	Kipf & Welling, 2017	Semi-supervised learning	Graph data	GCN	Foundation model for graph learning	Not domain-specific
3	Veličković et al., 2018	Attention mechanism	Graph data	GAT	Improved node importance learning	High computational cost
4	Hamilton et al., 2017	Inductive learning	Large graphs	GraphSAGE	Scalable GNN model	Sampling bias
5	Zitnik et al., 2018	Multimodal graph learning	Biomedical	GCN	Drug interaction prediction	Limited cancer focus
6	Yoon et al., 2024	Clinical graph modeling	EHR data	GCN	Improved diagnosis accuracy	Data dependency
7	Li et al., 2024	Multimodal fusion	Genomic + Clinical	Hybrid GNN	Enhanced subtype classification	Complexity
8	Chowa et al., 2023	Imaging-based graph	Histopathology	GNN-CNN	Improved tumor detection	Requires large datasets
9	Gogoshin et al., 2023	Oncology GNN survey	Multi-source	Multiple	Comprehensive oncology review	Limited benchmarking
10	Nair et al., 2022	Imaging graph analysis	MRI data	GCN	Improved classification	Limited generalization
11	Özcan et al., 2025	Heterogeneous graphs	Multi-omics	GNN	Improved feature integration	High dimensionality
12	Ju et al., 2024	Multi-omics integration	Genomic data	GAT	Captured complex dependencies	Computational overhead
13	Ramirez et	Graph classif	Biomedical	GCN	Cancer	Small dataset

	al., 2020	ication			classification	
14	Qureshi et al., 2024	Driver gene prediction	Genomic	GNN	Identified cancer genes	Limited clinical data
15	Ying et al., 2019	Explainable GNN	Graph data	GNN Explainer	Improved interpretability	Limited clinical use
16	Rieke et al., 2020	Federated learning	Healthcare data	FL + GNN	Privacy-preserving learning	Communication overhead
17	Oniani et al., 2021	Graph-based classification	Clinical	GCN	Improved prediction accuracy	Limited multimodal data
18	Boll et al., 2024	Scalable GNN	Large datasets	GraphSAGE	Improved scalability	Reduced accuracy tradeoff
19	Mesinovic et al., 2025	Hybrid deep learning	Multi-source	CNN + GNN	Improved performance	Model complexity
20	Zhang et al., 2022	Similarity learning	Clinical + Genomic	GNN	Improved patient clustering	Sensitive to noise

Models such as Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) consistently demonstrate superior performance in classification and prediction tasks due to their ability to leverage both feature information and graph topology. However, the comparison also reveals that most studies rely on static and predefined similarity measures, which may not adequately represent complex and dynamic relationships in heterogeneous biomedical data.

The comparative analysis presented in Table 1 highlights the evolving landscape of Graph Neural Network (GNN)-based approaches for cancer diagnosis using patient similarity networks. A clear trend observed across the reviewed studies is the shift from traditional machine learning models toward graph-based frameworks that explicitly capture inter-patient relationships. Models such as Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) consistently demonstrate superior performance in classification and prediction tasks due to their ability to leverage both feature information and graph topology. However, the comparison also reveals that most studies rely on static and predefined similarity measures, which may not adequately represent complex and dynamic relationships in heterogeneous biomedical data.

Another key observation is the increasing adoption of multimodal data integration, where clinical, genomic, and imaging data are combined to enhance diagnostic accuracy. While these approaches show improved performance, the integration strategies often lack standardization and fail to fully exploit cross-modal interactions. Additionally, despite the improvements in predictive accuracy, interpretability remains a significant limitation, as many models do not provide clinically meaningful explanations for their decisions.

From a computational perspective, scalability continues to be a challenge, particularly for large-scale patient datasets, as many models struggle with high memory and processing requirements. Furthermore, only a limited number of studies address data imbalance and noise, which are critical issues in real-world cancer datasets. Privacy-preserving techniques such as federated learning are emerging but are still underexplored in conjunction with GNN frameworks.

Overall, the analysis indicates that while GNN-based patient similarity networks have significantly advanced cancer diagnosis, there is a pressing need for adaptive graph construction methods, robust multimodal fusion techniques, scalable architectures, and explainable models. Addressing these gaps will be essential for transitioning from experimental research to reliable clinical deployment.

11. References

[1] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, 2019.

[2] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and Structural Biotechnology Journal*, vol. 13, pp. 8–17, 2015.

[3] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: A network-based approach to human disease," *Nature Reviews Genetics*, vol. 12, no. 1, pp. 56–68, 2011.

[4] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. ICLR*, 2017.

[5] S. Parisot, S. Ktena, E. Ferrante, M. Lee, B. Glocker, D. Rueckert, and S. Passerat-Palmbach, "Disease prediction using graph convolutional networks: Application to autism spectrum disorder and Alzheimer's disease," *Medical Image Analysis*, vol. 48, pp. 117–130, 2018.

[6] M. Zitnik, M. Agrawal, and J. Leskovec, "Modeling polypharmacy side effects with graph convolutional networks," *Bioinformatics*, vol. 34, no. 13, pp. i457–i466, 2018.

[7] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2021.

[8] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020.

[9] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. ICLR*, 2018.

[10] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Proc. NeurIPS*, pp. 1024–1034, 2017.

[11] G. Zhang, J. Li, Y. Wang, and H. Chen, "Patient similarity learning for disease prediction using graph-based approaches," *Scientific Reports*, vol. 12, pp. 14567, 2022.
[12] S. Ktena, S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert, "Metric learning with spectral graph convolutions on brain connectivity networks," *NeuroImage*, vol. 169, pp. 431–442, 2018.
[13] M. Zitnik and J. Leskovec, "Predicting multicellular function through multi-layer tissue networks," *Bioinformatics*, vol. 33, no. 14, pp. i190–i198, 2017.
[14] H. Cai, V. W. Zheng, and K. C. Chang, "A comprehensive survey of graph embedding: Problems, techniques, and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 9, pp. 1616–1637, 2018.
[15] N. Özcan Şimşek, A. Baştanlar, and M. Turan, "Heterogeneous graph neural networks for cancer classification using multi-omics data," *BMC Bioinformatics*, vol. 26, no. 1, pp. 1–15, 2025.
[16] H. J. Yoon, J. H. Lee, S. Kim, and D. Kim, "Graph-based deep learning for clinical decision support systems," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, pp. 1–12, 2024.
[17] B. Li, X. Wang, Y. Zhang, and Z. Li, "Multimodal graph neural networks for cancer subtype classification," *BMC Bioinformatics*, vol. 25, no. 1, pp. 1–14, 2024.
[18] R. Ramirez, J. A. Martinez, and L. Hernandez, "Cancer classification using graph convolutional networks," *Frontiers in Physics*, vol. 8, pp. 203, 2020.
[19] N. Rieke, J. Hancox, W. Li, F. Milletari, H. R. Roth, S. Albarqouni, S. Bakas, M. N. Galtier, B. A. Landman, and K. Maier-Hein, "The future of digital health with federated learning," *NPJ Digital Medicine*, vol. 3, no. 1, pp. 1–7, 2020.
[20] A. Nair, S. Kumar, R. Gupta, and P. Sharma, "Graph-based deep learning approaches for medical image analysis," *BMC Bioinformatics*, vol. 23, no. 1, pp. 1–12, 2022.
[21] R. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec, "GNNExplainer: Generating explanations for graph neural networks," in *Proc. NeurIPS*, 2019.
[22] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, and K. Bonawitz, "Advances and open problems in federated learning," *Foundations and Trends in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.
[23] G. Gogoshin, M. B. Sidorov, A. V. Ivanov, and A. A. Smirnov, "Graph neural networks in cancer research: Current trends and future directions," *Cancers*, vol. 15, no. 24, pp. 5858, 2023.
[24] H. Ju, Y. Zhao, X. Liu, and L. Wang, "Multi-omics data integration using graph neural networks for cancer analysis," *International Journal of Molecular Sciences*, vol. 25, no. 2, pp. 698, 2024.
[25] S. S. Chowa, R. Patel, M. Shah, and K. Mehta, "Graph neural network-based breast cancer diagnosis using histopathological images," *Journal of Cancer Research and Clinical Oncology*, vol. 149, no. 5, pp. 1–12, 2023.
[26] P. Sinthia, A. J. Albert, M. Malathi, G. Gurumoorthy, and S. Rajalakshmi, "Synergistic deep learning fusion for precision lung cancer staging," Asian Pacific Journal of Cancer Prevention, vol. 27, no. 6, pp. 2335–2343, 2026, doi: 10.31557/APJCP.2026.27.6.2335.
[27] S. B. Mohan, B. Ramesh, G. Gurumoorthy, and B. Kannan, "Brain tumour segmentation using deep learning method—Inception U-Net," International Journal of Medical Engineering and Informatics, vol. 18, no. 3, pp. 217–228, 2026, doi: 10.1504/IJMEI.2026.153152.