

An Explainable Patient-Similarity Graph Neural Network for Breast Cancer Recurrence and Survival Prognosis with Missing Multimodal Data

¹Sumayya Samreen

¹Assistant Professor, Department of Computer Science and Engineering, Methodist College of Engineering and Technology, Abids, Hyderabad, Telangana 500001, India

¹Email: sumayyasamreen01@gmail.com

ABSTRACT

Breast cancer prognosis requires the combined interpretation of clinical attributes, molecular profiles, and pathological findings, because no single data source fully describes the biological and clinical behaviour of a tumour. Multimodal deep learning has shown promise in integrating these sources, but two practical issues limit its adoption. First, real-world patient records are often incomplete, with one or more modalities missing for a considerable proportion of patients. Second, deep multimodal models are frequently difficult to interpret, which reduces clinical trust in their predictions. This paper proposes "An Explainable Patient-Similarity Graph Neural Network for Breast Cancer Recurrence and Survival Prognosis with Missing Multimodal Data". In the proposed framework, each patient is represented as a node in a population graph, and edges connect patients who are similar on the modalities that both of them have available. Modality-specific encoders generate masked embeddings that exclude missing data, and a graph attention network propagates information between similar patients, allowing a patient with a missing modality to benefit from the evidence of comparable patients rather than from artificial placeholder values. A multi-task prediction module jointly estimates recurrence status and a survival risk score, and an explanation module combines SHAP feature attributions, modality contribution analysis, and neighbour attention weights to show which features and which similar patients influenced each prediction. The framework can be evaluated using the concordance index, area under the receiver operating characteristic curve, F1-score, Kaplan–Meier risk stratification, and explanation consistency under varying missing-data rates. The proposed approach aims to provide robust and transparent prognostic support, although its predictions require independent clinical validation before use in medical decision-making.

Keywords: Breast cancer, prognosis, graph neural networks, patient similarity graph, missing data, multimodal learning, explainable artificial intelligence, recurrence prediction, survival analysis.

I. INTRODUCTION

Breast cancer is a leading cause of cancer incidence and mortality among women worldwide [1]. Advances in screening and therapy have improved outcomes, yet patients continue to experience a wide range of disease courses, from long-term remission to early recurrence and metastasis. Accurate prognosis allows clinicians to tailor adjuvant treatment and follow-up to each patient's expected risk. Lipkova et al. [2] reviewed the role of artificial intelligence in multimodal data integration for oncology and emphasised that combining clinical, molecular, radiological, and pathological information is essential for capturing the complexity of cancer, while also noting that

missing data and limited interpretability remain major barriers.

Most prognostic models treat each patient as an independent sample. In clinical reasoning, however, physicians often compare a new patient with previously treated patients who had similar characteristics. Graph-based learning offers a natural way to represent such relationships. Kipf and Welling [3] introduced graph convolutional networks, which learn node representations by aggregating information from neighbouring nodes, and Veličković et al. [4] proposed graph attention networks, which learn to weight the contribution of each neighbour. Parisot et al. [5] demonstrated that population graphs, in which patients are

nodes and edges encode phenotypic similarity, can be used with graph convolutional networks for disease prediction, showing that information sharing between similar patients improves classification.

Graph neural networks and multimodal fusion have also been applied to cancer. Wang et al. [6] proposed MOGONET, which constructs a patient similarity network for each omics type and uses graph convolutional networks to integrate multi-omics data for patient classification and biomarker identification. Wang et al. [7] developed GPDBN, a deep bilinear network that integrates genomic data and pathological images for breast cancer prognosis prediction by modelling both intra-modality and inter-modality interactions. Chen et al. [8] performed pan-cancer integrative histology-genomic analysis using multimodal deep learning and showed that the fusion of histology and genomics improves survival prediction across many cancer types.

Survival modelling is a central part of prognosis. Ching et al. [9] proposed Coxnet, a neural network extension of the Cox model for high-throughput omics data, and Lee et al. [10] introduced DeepHit, which learns the distribution of survival times directly and can handle competing risks. Huang et al. [11] developed SALMON, which combined multi-omics features and clinical data for breast cancer survival analysis. These approaches demonstrate that deep learning can model complex prognostic relationships, but they generally assume that every input feature is available for every patient.

Incomplete multimodal data remain a significant challenge. Vale-Silva and Rohr [12] showed that a multimodal survival model can be designed to accept missing modalities, and Cheerla and Gevaert [13] used multimodal dropout to make prognosis models more robust to absent data. Nevertheless, these approaches still make predictions for each patient in isolation. When a modality is missing, the model has less information about that patient and no mechanism to borrow evidence from similar patients. A patient-similarity graph provides

such a mechanism: if similarity is computed only from the modalities that two patients share, a patient missing gene expression data can still be connected to patients who are clinically and pathologically similar and who do have gene expression data.

Interpretability is equally important in clinical applications. Lundberg and Lee [14] proposed SHAP, a unified framework for interpreting model predictions based on Shapley values, which assigns each input feature a contribution to a particular prediction. Combining feature attributions with graph attention weights can explain not only which features influenced a prediction but also which comparable patients informed it, providing an explanation that resembles case-based clinical reasoning. Datasets such as METABRIC [15] provide clinical, gene expression, and copy number data with long-term follow-up for breast cancer patients, making them suitable for developing and evaluating such approaches.

To address these challenges, this paper proposes an Explainable Patient-Similarity Graph Neural Network for breast cancer recurrence and survival prognosis with missing multimodal data. The framework encodes each modality separately, builds a population graph using similarity computed over shared observed modalities, applies graph attention to propagate information between similar patients, and jointly predicts recurrence and survival risk. An explanation module presents feature-level, modality-level, and neighbour-level explanations for each prediction.

The primary objective of this research is to investigate whether structured information sharing between similar patients can reduce the performance loss caused by missing modalities while also making predictions more interpretable. The outputs of the framework are intended to support research and computer-aided prognostic assessment and should not replace professional clinical judgement.

II. LITERATURE SURVEY

Parisot et al. (2018) proposed the use of graph convolutional networks on population graphs for disease prediction, applying the

approach to autism spectrum disorder and Alzheimer's disease. Each subject was represented as a node with imaging-derived features, and edges were weighted using phenotypic information such as sex and acquisition site. The study demonstrated that exploiting relationships between subjects improves prediction compared with treating subjects independently. However, the graph construction assumed that the required phenotypic attributes were available for all subjects, and the approach was not designed for missing modalities.

Wang et al. (2021), MOGONET introduced a multi-omics integration framework that constructs a patient similarity network for each omics type and trains omics-specific graph convolutional networks. A view correlation discovery network then integrates the omics-specific predictions in a label space. The framework achieved strong performance in patient classification tasks and supported biomarker identification. However, it requires each patient to have all omics types available, and it focuses on classification rather than time-to-event survival prognosis.

Wang et al. (2021), GPDBN proposed a genomic and pathological deep bilinear network for breast cancer prognosis prediction. The model uses bilinear feature encoding to capture both intra-modality and inter-modality interactions between genomic data and pathological image features. The study showed that modelling cross-modal interactions improves breast cancer prognosis compared with simple concatenation. However, the bilinear fusion requires both modalities for every patient, and the model provides limited explanation of individual predictions.

Lee et al. (2018) introduced DeepHit, a deep learning approach to survival analysis that directly learns the distribution of first hitting times without assuming a proportional hazards form. DeepHit can model competing risks and uses a combination of log-likelihood and ranking losses. The model demonstrated improved discrimination compared with conventional survival models. However, it operates on a single feature vector and does not address

multimodal integration or missing modalities.

Huang et al. (2019) developed SALMON, a survival analysis model for breast cancer that uses co-expression network analysis to reduce the dimensionality of mRNA and microRNA data and combines the resulting modules with copy number, mutation, and clinical information in a Cox-based neural network. The approach improved prognostic performance and provided biological interpretability through gene modules. However, it assumes complete multi-omics profiles and does not share information between related patients.

Lundberg and Lee (2017) proposed SHapley Additive exPlanations (SHAP), a unified approach for interpreting model predictions based on game-theoretic Shapley values. SHAP assigns each feature an importance value for a specific prediction and satisfies desirable properties such as local accuracy and consistency. It has been widely adopted for explaining clinical prediction models. However, feature attribution alone does not explain relational evidence, such as the influence of similar patients in a graph-based model, which motivates combining SHAP with attention-based neighbour explanations in the proposed framework.

III. EXISTING SYSTEM

The existing systems for breast cancer prognosis mainly rely on clinical staging and statistical survival models, as well as machine learning and deep learning models that combine clinical attributes with genomic or pathological features. Multimodal deep learning systems typically extract features from each modality, fuse them through concatenation, attention, or bilinear pooling, and predict outcomes such as survival risk or recurrence. These systems treat each patient as an independent sample and learn a mapping from that patient's features to the outcome.

These approaches face two main limitations. First, they generally require complete multimodal data. When a modality is missing, patients are excluded or missing values are replaced by zeros or averages,

which reduces the effective dataset size and can mislead the model. Because each patient is modelled independently, a patient with missing data has less information available, and there is no mechanism to draw on the evidence of comparable patients. Second, most multimodal deep learning models function as black boxes. They provide a risk score without indicating which features or modalities contributed to it, which limits clinicians' ability to verify predictions and reduces trust. Therefore, existing systems do not adequately combine robustness to missing data with transparent, clinically meaningful explanations.

IV. PROPOSED SYSTEM

The proposed system introduces an Explainable Patient-Similarity Graph Neural Network for breast cancer recurrence and survival prognosis with missing multimodal data. The framework uses clinical attributes, genomic profiles, and pathology-derived features. Each modality is processed by its own encoder, and an availability mask ensures that missing modalities do not contribute artificial values. The masked embeddings are combined into a node representation for each patient.

A patient-similarity graph is then constructed in which each patient is a node. The similarity between two patients is calculated only over the modalities that both patients have available and is normalised by the number of shared modalities, so that missing data do not create false similarity or dissimilarity. Each patient is connected to its k most similar patients. A graph attention network learns how strongly each neighbour should influence a patient's representation, allowing patients with missing modalities to receive complementary information from similar patients. The resulting node representations are passed to two task-specific heads: one predicts recurrence status, and the other estimates a survival risk score using the Cox partial likelihood. An explanation module provides SHAP-based feature attributions, modality contribution scores, and the attention weights assigned to neighbouring patients. The framework can be evaluated using the concordance index, AUC-ROC, F1-score, Kaplan–Meier risk

stratification, and explanation consistency across different missing-data rates. The predictions are intended to support research and computer-aided prognosis rather than replace clinical decision-making.

V. SYSTEM ARCHITECTURE

The system architecture of the proposed Explainable Patient-Similarity Graph Neural Network consists of several interconnected modules that learn prognostic representations from incomplete multimodal patient data and explain the resulting predictions.

The first module is the Patient Data module, which collects clinical attributes, genomic profiles, and pathology-derived features for each patient, together with recurrence status, survival time, and event indicator. Clinical attributes may include age, tumour size, grade, lymph node status, and receptor status. Genomic profiles may include gene expression or copy number data, while pathology features may be extracted from digitised slides or pathology reports. Some modalities are missing for some patients, and an availability mask records this information.

The second module is Feature Encoding. Each modality is preprocessed and encoded using a dedicated neural network, such as a multilayer perceptron for clinical and pathology features and a dimensionality-reducing encoder for high-dimensional genomic data. The encoders produce embeddings of the same dimension. Embeddings of missing modalities are masked out rather than replaced with zeros, and the node feature is computed as the average of the available modality embeddings.

The third module is the Patient Similarity Graph. Similarity between two patients is computed using cosine similarity over the embeddings of the modalities they share and is averaged across shared modalities. Each patient is connected to the k patients with the highest similarity, and an edge is created only when the two patients share at least one modality. In the transductive setting, training, validation, and test patients are included in the same graph, but only the

labels of training patients are used during learning.

The fourth module is the Graph Neural Network. Two or more graph attention layers update each node representation by aggregating information from its neighbours, with learned attention coefficients that reflect the relevance of each neighbour. Through this message passing, a patient with a missing modality can incorporate evidence from similar patients who have that modality available.

The fifth module is the Multi-Task Heads module. The final node representation is passed to a recurrence prediction head with a sigmoid output and to a survival head that produces a log-risk score. The two tasks are trained jointly, allowing the shared representation to capture information that is useful for both short-term recurrence and long-term survival.

The sixth module is the Explanation Module. SHAP values are computed for the input features to identify the clinical and molecular variables that most strongly influenced a prediction. Modality contribution scores are calculated by comparing predictions with and without each available modality. Graph attention weights identify the neighbouring patients that most influenced the prediction, providing a case-based explanation.

Finally, the Evaluation module assesses the framework using the concordance index for survival, AUC-ROC and F1-score for recurrence, Kaplan–Meier curves with log-rank tests for risk stratification, and the stability of explanations across repeated runs. Experiments with different missing-modality rates and comparisons with non-graph multimodal models and imputation-based baselines can indicate the benefit of patient-level information sharing.

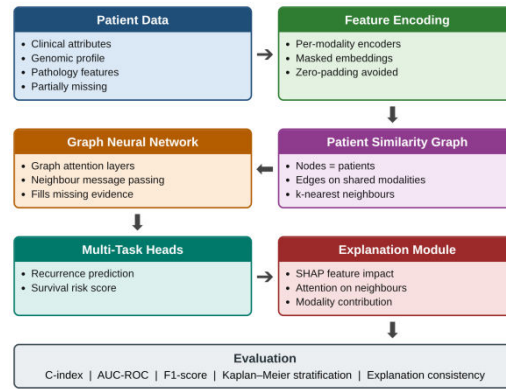


Fig 5.1 System Architecture

VI. ALGORITHM

Algorithm: Explainable Patient-Similarity Graph Neural Network for Breast Cancer Prognosis (EPSG-BCP)

Input:

$D \leftarrow$ Breast cancer cohort with clinical, genomic, and pathology modalities, recurrence labels y^{rec} , survival time T , and event indicator δ .

$m_{ik} \leftarrow$ Availability mask, equal to 1 if modality k is observed for patient i and 0 otherwise.

$k_n \leftarrow$ Number of nearest neighbours in the patient graph.

$L_g \leftarrow$ Number of graph attention layers.

$\gamma \leftarrow$ Weight balancing the recurrence and survival losses.

$E_{\text{max}} \leftarrow$ Maximum number of training epochs.

Output:

$M^* \leftarrow$ Trained graph-based prognosis model.

$\hat{y}^{\text{rec}}, \hat{r} \leftarrow$ Predicted recurrence probability and survival risk score.

$X \leftarrow$ Feature, modality, and neighbour explanations for each prediction.

Steps:

1. Start.
2. Load the cohort D and construct the availability mask m_{ik} for every patient and modality.

3. Divide the patients into training, validation, and testing subsets using a stratified split based on recurrence and event status.

4. Preprocess each modality using statistics from the training subset, including encoding, normalisation, and dimensionality reduction of genomic features.

5. Compute the embedding e_{ik} of each observed modality using its encoder, and compute the node feature as the average of the available embeddings:

$$h_i^{(0)} = \sum_k m_{ik} e_{ik} / \sum_k m_{ik}$$

6. For every pair of patients i and j , compute similarity over the shared observed modalities:

$$s_{ij} = \sum_k m_{ik} m_{jk} \cos(e_{ik}, e_{jk}) / \sum_k m_{ik} m_{jk}$$

7. Connect each patient to the k_n patients with the highest similarity, considering only pairs that share at least one modality.

8. Apply L_g graph attention layers. For each layer, compute the attention coefficient between patient i and neighbour j and update the representation:

$$\alpha_{ij} = \text{softmax}_j(\text{LeakyReLU}(a^T [W h_i \parallel W h_j]))$$

$$h_i' = \sigma(\sum_{j \in N(i)} \alpha_{ij} W h_j)$$

9. Pass the final representation to the recurrence head to obtain $\hat{y}^{\text{rec}}$ and to the survival head to obtain the log-risk score $\hat{r}$.

10. Calculate the binary cross-entropy loss for recurrence and the Cox negative partial log-likelihood for survival using training patients only, and combine them:

$$L = L_{\text{Cox}} + \gamma \cdot L_{\text{BCE}}$$

11. Update all encoder, graph, and prediction parameters using the Adam optimiser, applying dropout on edges and features for regularisation.

12. Evaluate the concordance index and AUC-ROC on the validation patients after each epoch, and apply early stopping when performance stops improving.

13. Save the best model M^* .

14. Evaluate M^* on the test patients, calculating the concordance index, AUC-

ROC, and F1-score, and stratify patients into risk groups for Kaplan–Meier and log-rank analysis.

15. Repeat the evaluation after removing modalities at increasing rates to measure robustness to missing data.

16. For each test prediction, compute SHAP values for the input features, modality contribution scores by removing each available modality in turn, and the top neighbours ranked by attention weight α_{ij} .

17. For a new patient, encode the available modalities, connect the patient to its most similar patients in the existing graph, and compute the predictions and explanations.

18. Return the recurrence probability, survival risk score, risk group, and explanations X together with the evaluation results.

19. Stop.

Computational Complexity

Let N be the number of patients, K the number of modalities, d the embedding dimension, and k_n the number of neighbours. Computing pairwise similarities requires $O(N^2 K d)$ operations, which can be reduced using approximate nearest-neighbour search. Each graph attention layer requires $O(N k_n d + N d^2)$ operations, so training over E epochs with L_g layers has an approximate complexity of

$$O(N^2 K d + E L_g (N k_n d + N d^2))$$

Because the graph is sparse and breast cancer cohorts typically contain a few thousand patients, the framework remains computationally feasible on standard hardware.

VII. CONCLUSION

The proposed Explainable Patient-Similarity Graph Neural Network addresses two important limitations of multimodal breast cancer prognosis models: sensitivity to missing modalities and limited interpretability. By representing patients as nodes in a population graph and computing similarity only over shared observed modalities, the framework allows patients with incomplete data to benefit from the evidence of comparable patients rather than

from artificial placeholder values. Graph attention layers learn how strongly each neighbour should influence a prediction, and a multi-task design jointly predicts recurrence and survival risk from a shared representation. The explanation module combines SHAP feature attributions, modality contribution scores, and neighbour attention to provide feature-level, modality-level, and case-based explanations. The effectiveness of the framework can be assessed using the concordance index, AUC-ROC, F1-score, Kaplan–Meier stratification, and explanation stability under different missing-data rates. Overall, the framework provides a promising research direction for robust and transparent breast cancer prognosis, although rigorous experimental evaluation and independent clinical validation are required before practical use.

VIII. FUTURE SCOPE

The future scope of the proposed framework includes incorporating richer modalities, such as whole-slide histopathology images, mammography, MRI, and longitudinal treatment records, into the patient graph. Inductive graph learning methods can be investigated so that new patients can be added without retraining the entire model. Heterogeneous graphs that include genes, pathways, and treatments as additional node types could provide more biologically meaningful explanations. Federated graph learning could enable collaboration between hospitals while keeping patient data within each institution. Further work should also examine privacy considerations, because neighbour explanations refer to other patients and must be presented in an anonymised form. Finally, external validation on multi-institutional cohorts, fairness assessment across demographic groups, and user studies with clinicians will be necessary to determine whether the explanations are clinically useful and whether the framework is suitable for decision support.

REFERENCES

[1] F. Bray et al., "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for*

Clinicians, vol. 74, no. 3, pp. 229–263, 2024. doi: 10.3322/caac.21834.

[2] J. Lipkova et al., "Artificial intelligence for multimodal data integration in oncology," *Cancer Cell*, vol. 40, no. 10, pp. 1095–1110, 2022. doi: 10.1016/j.ccell.2022.09.012.

[3] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *Proc. International Conference on Learning Representations (ICLR)*, 2017.

[4] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *Proc. International Conference on Learning Representations (ICLR)*, 2018.

[5] S. Parisot et al., "Disease prediction using graph convolutional networks: Application to autism spectrum disorder and Alzheimer's disease," *Medical Image Analysis*, vol. 48, pp. 117–130, 2018. doi: 10.1016/j.media.2018.06.002.

[6] T. Wang et al., "MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification," *Nature Communications*, vol. 12, Art. no. 3445, 2021. doi: 10.1038/s41467-021-23774-w.

[7] Z. Wang, R. Li, M. Wang, and A. Li, "GPDBN: Deep bilinear network integrating both genomic data and pathological images for breast cancer prognosis prediction," *Bioinformatics*, vol. 37, no. 18, pp. 2963–2970, 2021. doi: 10.1093/bioinformatics/btab185.

[8] R. J. Chen et al., "Pan-cancer integrative histology-genomic analysis via multimodal deep learning," *Cancer Cell*, vol. 40, no. 8, pp. 865–878, 2022. doi: 10.1016/j.ccell.2022.07.004.

[9] T. Ching, X. Zhu, and L. X. Garmire, "Cox-nnet: An artificial neural network method for prognosis prediction of high-throughput omics data," *PLOS Computational Biology*, vol. 14, no. 4, Art. no. e1006076, 2018. doi: 10.1371/journal.pcbi.1006076.

- [10] C. Lee, W. Zame, J. Yoon, and M. van der Schaar, "DeepHit: A deep learning approach to survival analysis with competing risks," in Proc. AAAI Conference on Artificial Intelligence, vol. 32, no. 1, 2018. doi: 10.1609/aaai.v32i1.11842.
- [11] Z. Huang et al., "SALMON: Survival analysis learning with multi-omics neural networks on breast cancer," *Frontiers in Genetics*, vol. 10, Art. no. 166, 2019. doi: 10.3389/fgene.2019.00166.
- [12] L. A. Vale-Silva and K. Rohr, "Long-term cancer survival prediction using multimodal deep learning," *Scientific Reports*, vol. 11, Art. no. 13505, 2021. doi: 10.1038/s41598-021-92799-4.
- [13] A. Cheerla and O. Gevaert, "Deep learning with multimodal representation for pancancer prognosis prediction," *Bioinformatics*, vol. 35, no. 14, pp. i446–i454, 2019. doi: 10.1093/bioinformatics/btz342.
- [14] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, pp. 4765–4774, 2017.
- [15] C. Curtis et al., "The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups," *Nature*, vol. 486, pp. 346–352, 2012. doi: 10.1038/nature10983.