
**A Missing-Modality Robust Attention Fusion Network for Breast Cancer
Survival Prediction Using Incomplete Clinical, Genomic and Histopathology
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 is one of the most frequently diagnosed cancers worldwide, and accurate prognosis is essential for planning treatment intensity, follow-up schedules, and patient counselling. The survival outcome of a patient is influenced by many complementary factors, including clinical and pathological attributes, molecular alterations captured by gene expression, and the tissue morphology visible in histopathology images. Multimodal deep learning can combine these heterogeneous sources to produce more informative prognostic estimates than any single modality. However, in real hospital and research cohorts, many patients do not have every modality available because of cost, tissue availability, differences in clinical protocols, or incomplete records. Most existing multimodal models assume complete data and therefore either discard incomplete patients or fill the missing values with simple placeholders, which reduces the size of the usable cohort and may introduce bias. This paper proposes "A Missing-Modality Robust Attention Fusion Network for Breast Cancer Survival Prediction Using Incomplete Clinical, Genomic and Histopathology Data" to address this limitation. The framework uses modality-specific encoders, a learnable null representation for absent modalities, a masked cross-modal attention fusion mechanism, and modality dropout during training so that the model learns to make reliable predictions from any available subset of modalities. A Cox proportional hazards objective is used to estimate an individual risk score, which can then be used to stratify patients into risk groups and generate survival curves. The performance of the framework can be evaluated using the concordance index, time-dependent area under the curve, integrated Brier score, and Kaplan–Meier log-rank analysis under different missing-modality rates. The proposed approach aims to make multimodal prognosis practical for incomplete real-world data, although its predictions require independent clinical validation before use in medical decision-making.

Keywords: Breast cancer, survival prediction, multimodal deep learning, missing modalities, attention fusion, modality dropout, Cox proportional hazards, histopathology, gene expression.

I. INTRODUCTION

Breast cancer is the most commonly diagnosed cancer among women and remains a leading cause of cancer-related death across the world [1]. Patients with the same clinical stage can experience very different outcomes, because tumour behaviour depends on a combination of clinical, molecular, and morphological factors. Prognosis, which refers to estimating the likely course of the disease and the time until events such as death or recurrence, therefore

plays a central role in deciding the intensity of treatment and the frequency of follow-up. Conventional prognostic tools rely mainly on tumour size, lymph node status, histological grade, and hormone receptor status. Although these indicators remain clinically valuable, they do not fully capture the biological heterogeneity of breast cancer.

Large-scale studies have shown that breast tumours contain distinct molecular subgroups with different clinical outcomes. Curtis et al. [2] analysed the genomic and transcriptomic architecture of approximately

2,000 breast tumours in the METABRIC cohort and identified novel subgroups with distinct survival patterns, demonstrating that molecular data provide prognostic information beyond conventional clinical variables. At the same time, histopathology slides contain rich information about tumour architecture, nuclear morphology, and the immune microenvironment. Boehm et al. [3] highlighted that integrating such complementary data types, including clinical, molecular, radiological, and pathological data, is an important direction for advancing precision oncology.

Deep learning has made it possible to learn predictive representations directly from these heterogeneous data sources. Sun et al. [4] developed a multimodal deep neural network that integrated gene expression, copy number alteration, and clinical data for breast cancer prognosis prediction, and showed that the combined model outperformed single-modality models. Cheerla and Gevaert [5] proposed a multimodal representation learning approach for pan-cancer prognosis prediction that combined clinical data, gene expression, microRNA, and whole-slide images. Mobadersany et al. [6] demonstrated that convolutional networks trained on histology images, together with genomic markers, could predict cancer outcomes. Chen et al. [7] further introduced Pathomic Fusion, an integrated framework that fused histopathology and genomic features using tensor-based fusion and gating-based attention for cancer diagnosis and prognosis.

Survival prediction differs from ordinary classification because many patients are censored; that is, the event of interest has not been observed by the end of follow-up. Katzman et al. [8] proposed DeepSurv, a deep neural network trained with the Cox partial likelihood, which allows non-linear relationships between patient features and risk to be modelled while correctly handling censored observations. Such survival-oriented objectives are an important building block for multimodal prognosis models.

A major practical obstacle is that multimodal data are frequently incomplete. In public

cohorts and hospital records, a patient may have clinical data but no gene expression profile, or gene expression but no digitised slide. Vale-Silva and Rohr [9] developed MultiSurv, a long-term multimodal survival prediction model that was designed to tolerate missing input modalities. Ma et al. [10] investigated multimodal learning when a large proportion of samples lack one modality, and Ma et al. [11] showed that multimodal transformers can be highly sensitive to missing modalities and that the fusion strategy strongly affects robustness. Earlier, Havaei et al. [12] proposed hetero-modal learning that computes statistics over whichever modalities are present, demonstrating that architectures can be designed to accept a variable set of inputs.

Attention-based fusion has also emerged as an effective way to model the interactions between modalities. Chen et al. [13] proposed a multimodal co-attention transformer in which genomic features guide attention over histology patches for survival prediction. Arya and Saha [14] investigated multimodal deep learning architectures for breast cancer survival prediction using clinical and genomic data, while Steyaert et al. [15] reviewed multimodal data fusion for cancer biomarker discovery and identified missing data as one of the most important open challenges. These studies show that multimodal fusion is promising, but that robustness to incomplete data must be addressed directly in the model design rather than treated as a preprocessing problem.

To address these challenges, this paper proposes a Missing-Modality Robust Attention Fusion Network for breast cancer survival prediction. The framework accepts clinical, genomic, and histopathology data, together with a binary availability mask that records which modalities are present for each patient. Each modality is processed by its own encoder, and missing modalities are represented by a learnable null token rather than artificial values. A masked cross-modal attention module fuses only the available modality embeddings, and modality dropout is applied during training to expose the model to many combinations of missing

inputs. The fused representation is passed to a Cox-based survival head that outputs a patient-specific risk score.

The primary objective of this research is to investigate whether a single multimodal model can provide reliable survival predictions for patients with complete data as well as for patients with one or more missing modalities, without discarding incomplete records. The outputs of the framework are intended to support research and computer-aided prognostic assessment and should not be considered a substitute for clinical judgement.

II. LITERATURE SURVEY

Sun et al. (2019) proposed a multimodal deep neural network, known as MDNNMD, for human breast cancer prognosis prediction by integrating gene expression profiles, copy number alteration data, and clinical information. Separate deep networks were trained for each data type and their outputs were combined at the score level. The study demonstrated that multi-dimensional data improve prognostic performance compared with single data types. However, the approach assumes that every patient has all data types available, and it does not provide a mechanism for handling incomplete modalities.

Cheerla and Gevaert (2019) developed a deep learning model with multimodal representation for pan-cancer prognosis prediction using clinical data, mRNA expression, microRNA expression, and whole-slide images. Their method learned a shared representation using a similarity loss and applied multimodal dropout so that the network became less dependent on any single modality. The work showed that representation learning can make multimodal models more robust. However, the evaluation focused on pan-cancer prognosis, and the specific characteristics of breast cancer cohorts were not studied in detail.

Katzman et al. (2018) introduced DeepSurv, a Cox proportional hazards deep neural network that models non-linear relationships between covariates and the risk of an event.

DeepSurv correctly handles censored observations through the negative partial log-likelihood and can provide personalised treatment recommendations. The model is a widely used foundation for deep survival analysis. However, it was designed for a single feature vector and does not address the integration of heterogeneous or incomplete modalities.

Chen et al. (2022) proposed Pathomic Fusion, an interpretable framework that combines histopathology image features and genomic features using gating-based attention and Kronecker product fusion for cancer diagnosis and prognosis. The framework demonstrated improved survival prediction compared with unimodal models and provided interpretable visualisations. However, tensor-based fusion requires all modalities at inference time and can produce very high-dimensional representations, which increases computational cost.

Vale-Silva and Rohr (2021) developed MultiSurv, a multimodal deep learning model for long-term cancer survival prediction that integrates clinical, genomic, and imaging data. The model uses dedicated submodels for each data type and a fusion layer that can operate when some modalities are missing. The work demonstrated that incomplete multimodal data can still be exploited for survival prediction. However, the fusion strategy treats modalities largely independently and does not explicitly model cross-modal interactions.

Ma et al. (2022) investigated whether multimodal transformers are robust to missing modalities and found that their performance can degrade sharply when one modality is absent at test time. They showed that the optimal fusion strategy depends on the dataset and proposed an approach for automatically searching for robust fusion configurations. This study highlights that robustness to missing modalities must be explicitly considered during model design, which motivates the masked attention and modality dropout strategies used in the proposed framework.

III. EXISTING SYSTEM

The existing systems for breast cancer prognosis mainly rely on clinical and pathological indicators, such as tumour size, lymph node involvement, histological grade, and hormone receptor status, which are combined using statistical models such as the Cox proportional hazards model or clinical scoring tools. More recent research uses machine learning and deep learning models that combine clinical attributes with gene expression data or histopathology images. Multimodal deep learning models typically extract features from each modality and combine them through concatenation, score averaging, or tensor fusion before predicting a survival outcome.

Most of these multimodal systems assume that every patient has a complete set of modalities. When data are missing, the common practice is to remove incomplete patients from the cohort, to replace missing feature vectors with zeros, or to use simple statistical imputation such as mean values. Removing patients reduces the training sample size and may bias the cohort towards patients who received more extensive diagnostic testing. Zero-filling and mean imputation can mislead the model, because an artificial value cannot be distinguished from a real measurement. In addition, models trained only on complete data often perform poorly when a modality is unavailable at the time of prediction. Therefore, existing systems do not provide an ideal balance between multimodal predictive power and robustness to the incomplete data found in real clinical settings.

IV. PROPOSED SYSTEM

The proposed system introduces a Missing-Modality Robust Attention Fusion Network for breast cancer survival prediction using incomplete clinical, genomic, and histopathology data. The framework is designed to use all available patient records, including those with one or more missing modalities, rather than restricting training to patients with complete data. Each modality is encoded by a dedicated neural network suited

to its data type: a multilayer perceptron for clinical variables, a self-normalising network for high-dimensional gene expression data, and a convolutional neural network with patch aggregation for histopathology whole-slide images. A binary availability mask indicates which modalities are present for each patient, and absent modalities are replaced by a learnable null token so that the model can explicitly recognise missing information.

The modality embeddings are fused using masked cross-modal attention, in which attention weights are computed only over the modalities that are available. This allows each available modality to draw complementary information from the others while ignoring absent inputs. During training, modality dropout randomly hides observed modalities, which simulates missing-data patterns and prevents the network from relying too heavily on any single source. The fused representation is passed to a survival prediction head trained with the Cox negative partial log-likelihood, producing a risk score for each patient. Patients can then be stratified into low-risk and high-risk groups, and survival curves can be estimated for each group. The framework can be evaluated using the concordance index, time-dependent AUC, integrated Brier score, and log-rank tests, both on complete data and under controlled missing-modality rates. The resulting predictions are intended to support research and computer-aided prognosis rather than replace professional medical assessment.

V. SYSTEM ARCHITECTURE

The system architecture of the proposed Missing-Modality Robust Attention Fusion Network consists of several interconnected modules designed to estimate breast cancer survival risk from whichever modalities are available for a patient.

The first module is the Multimodal Input module, which collects clinical records, gene expression profiles, and histopathology whole-slide images for each patient, together with survival time and event status. Clinical

records may include age at diagnosis, tumour size, lymph node status, histological grade, and hormone receptor status. A binary availability mask is created for every patient to record whether each modality is present or missing. Publicly available cohorts such as METABRIC and TCGA-BRCA can be used, depending on the modalities required.

The second module is Preprocessing, which prepares each modality for learning. Clinical variables are cleaned, categorical attributes are encoded, and continuous attributes are normalised. Gene expression data are log-transformed and standardised, and a subset of informative genes can be selected using variance filtering or univariate survival association computed only on the training data. Histopathology slides are tiled into fixed-size patches, background regions are removed, and stain normalisation can be applied to reduce colour variations across laboratories.

The third module is the Modality Encoders module. A multilayer perceptron encodes clinical variables, a self-normalising neural network encodes gene expression features, and a pretrained convolutional neural network extracts patch-level features from histopathology images, which are aggregated into a slide-level embedding using attention-based pooling. All encoders project their outputs into embeddings of the same dimension. If a modality is missing, its embedding is replaced by a learnable null token.

The fourth module is Modality Dropout, which is active only during training. For each training sample, observed modalities are randomly hidden with a fixed probability, while ensuring that at least one modality remains available. This exposes the model to a wide range of missing-data combinations and encourages each modality encoder to learn independently useful features.

The fifth module is Masked Attention Fusion. The modality embeddings are treated as a short sequence of tokens and passed through a multi-head cross-modal attention layer. The availability mask prevents

attention to missing modalities, so that only real information contributes to the fused representation. A gating layer then weights the contribution of each modality, and the gated embeddings are pooled into a single patient representation.

The sixth module is Survival Prediction. The fused representation is passed through fully connected layers to produce a scalar log-risk score. The model is trained using the Cox negative partial log-likelihood, which correctly handles censored patients. The predicted risk score is used to divide patients into risk groups using the median risk of the training set, and survival curves can be estimated for each group.

Finally, the Evaluation module assesses the framework using the concordance index, time-dependent AUC, integrated Brier score, and the Kaplan–Meier log-rank test. The model is evaluated on the complete test set as well as on test sets where modalities are artificially removed at different rates, which measures robustness to missing data. Comparisons with unimodal models, complete-case multimodal models, and imputation-based baselines can indicate the benefit of the proposed design.

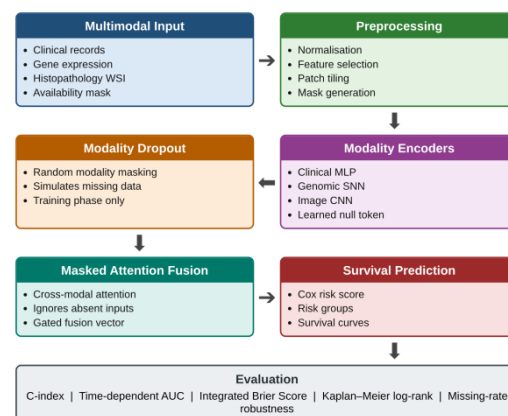


Fig 5.1 System Architecture

VI. ALGORITHM

Algorithm: Missing-Modality Robust Attention Fusion Network for Breast Cancer Survival Prediction (MRAF-Surv)

Input:

$D \leftarrow$ Breast cancer cohort with clinical data X_c , gene expression data X_g , histopathology images X_h , survival time T , and event indicator δ .

$m \leftarrow$ Availability mask, where $m_k = 1$ if modality k is present and $m_k = 0$ otherwise.

$p \leftarrow$ Modality dropout probability.

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

$B \leftarrow$ Batch size.

$\eta \leftarrow$ Learning rate.

Output:

$M^* \leftarrow$ Trained survival prediction model.

$\hat{r} \leftarrow$ Predicted risk score for each patient.

$P \leftarrow$ Survival prediction and robustness metrics.

Steps:

1. Start.
2. Load the breast cancer cohort D and construct the availability mask m for each patient.
3. Remove records without valid survival time or event status, and divide the patients into training, validation, and testing subsets using a stratified split based on event status.
4. Preprocess each modality using statistics computed only from the training subset:
 - Encode categorical clinical attributes and normalise continuous attributes.
 - Log-transform and standardise gene expression values, and select informative genes.
 - Tile histopathology slides into patches, remove background, and apply stain normalisation.
5. Initialise the clinical encoder f_c , genomic encoder f_g , and image encoder f_h , together with the learnable null token e_{null} .
6. For each training epoch and each mini-batch of size B , apply modality dropout by setting $m_k = 0$ for observed modalities with

probability p , keeping at least one modality available.

7. Compute the embedding of each modality:

$$e_k = m_k \cdot f_k(x_k) + (1 - m_k) \cdot e_{\text{null}}$$

8. Apply masked multi-head attention over the embeddings, assigning zero attention weight to modalities with $m_k = 0$:

$$a_{kj} = \text{softmax}_j(q_k \cdot k_j / \sqrt{d} + \log m_j)$$

9. Compute gated modality weights and obtain the fused patient representation z by pooling the gated embeddings.

10. Compute the log-risk score $r = g(z)$ using the survival prediction head.

11. Calculate the Cox negative partial log-likelihood loss:

$$L = - \sum_{i: \delta_i = 1} [r_i - \log \sum_{j: T_j \geq T_i} \exp(r_j)]$$

12. Update the model parameters using the Adam optimiser with learning rate η , and apply weight decay for regularisation.

13. Evaluate the concordance index on the validation subset after each epoch, and apply early stopping if it does not improve for the specified patience period.

14. Save the model M^* that achieves the best validation concordance index.

15. Evaluate M^* on the independent test subset, both with all available modalities and with modalities artificially removed at different missing rates.

16. Calculate the concordance index, time-dependent AUC, and integrated Brier score.

17. Divide test patients into low-risk and high-risk groups using the median training risk score, and compare their survival curves using the log-rank test.

18. For a new patient, apply the same preprocessing, construct the availability mask, and compute the predicted risk score $\hat{r}$.

19. Return the risk score, risk group, and estimated survival curve together with the evaluation results.

20. Stop.

Computational Complexity

Let N denote the number of training patients, E the number of epochs, K the number of modalities, and F the cost of encoding one patient across all available modalities. Because attention is computed over only K modality tokens, the fusion cost per patient is $O(K^2d)$, where d is the embedding dimension. The approximate training complexity is

$$O(NE(F + K^2d))$$

Since K is small (three modalities in this work), the fusion cost is negligible compared with the encoding cost, which is dominated by histopathology feature extraction. Precomputing patch features can further reduce training time.

VII. CONCLUSION

The proposed Missing-Modality Robust Attention Fusion Network for breast cancer survival prediction addresses one of the most important practical limitations of multimodal prognosis models: the frequent absence of one or more data modalities. The framework integrates clinical, genomic, and histopathology data through modality-specific encoders, represents missing modalities with a learnable null token, and fuses only the available information through masked cross-modal attention. Modality dropout during training exposes the model to realistic missing-data patterns, enabling it to produce consistent risk estimates for patients with complete as well as incomplete records. The Cox-based survival head correctly handles censored observations and produces individual risk scores that can be used for risk stratification and survival curve estimation. The effectiveness of the framework can be assessed using the concordance index, time-dependent AUC, integrated Brier score, and log-rank analysis under different missing-modality rates. Overall, the framework provides a practical research direction for using incomplete real-world multimodal data in breast cancer prognosis, although its clinical reliability must be established through rigorous experimental testing and independent validation.

VIII. FUTURE SCOPE

The future scope of the proposed framework includes extending it to additional modalities, such as mammography, breast MRI, and other molecular data types, including DNA methylation and copy number alterations. Evaluation on larger, multi-institutional cohorts can improve robustness across different laboratories, scanners, and patient populations. Explainable artificial intelligence techniques, such as attention visualisation on histopathology patches and feature attribution for clinical and genomic variables, can be integrated to make predictions more transparent to clinicians. Federated learning may allow the model to be trained across hospitals without centralising sensitive patient data. The framework could also be extended to predict multiple endpoints, such as recurrence-free survival and distant metastasis, and to model competing risks. Finally, prospective validation, fairness assessment across demographic groups, and clinical usability studies will be essential before any practical deployment.

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] 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.
- [3] K. M. Boehm, P. Khosravi, R. Vanguri, J. Gao, and S. P. Shah, "Harnessing multimodal data integration to advance precision oncology," *Nature Reviews Cancer*, vol. 22, pp. 114–126, 2022. doi: 10.1038/s41568-021-00408-3.
- [4] D. Sun, M. Wang, and A. Li, "A multimodal deep neural network for human breast cancer prognosis prediction by

- integrating multi-dimensional data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 3, pp. 841–850, 2019. doi: 10.1109/TCBB.2018.2806438.
- [5] 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.
- [6] P. Mobadersany et al., "Predicting cancer outcomes from histology and genomics using convolutional networks," *Proceedings of the National Academy of Sciences*, vol. 115, no. 13, pp. E2970–E2979, 2018. doi: 10.1073/pnas.1717139115.
- [7] R. J. Chen et al., "Pathomic Fusion: An integrated framework for fusing histopathology and genomic features for cancer diagnosis and prognosis," *IEEE Transactions on Medical Imaging*, vol. 41, no. 4, pp. 757–770, 2022. doi: 10.1109/TMI.2020.3021387.
- [8] J. L. Katzman et al., "DeepSurv: Personalized treatment recommender system using a Cox proportional hazards deep neural network," *BMC Medical Research Methodology*, vol. 18, Art. no. 24, 2018. doi: 10.1186/s12874-018-0482-1.
- [9] 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.
- [10] M. Ma et al., "SMIL: Multimodal learning with severely missing modality," in *Proc. AAAI Conference on Artificial Intelligence*, vol. 35, no. 3, pp. 2302–2310, 2021. doi: 10.1609/aaai.v35i3.16330.
- [11] M. Ma, J. Ren, L. Zhao, D. Testuggine, and X. Peng, "Are multimodal transformers robust to missing modality?," in *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18177–18186, 2022.
- [12] M. Havaei, N. Guizard, N. Chapados, and Y. Bengio, "HeMIS: Hetero-modal image segmentation," in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS vol. 9901, pp. 469–477, 2016. doi: 10.1007/978-3-319-46723-8_54.
- [13] R. J. Chen et al., "Multimodal co-attention transformer for survival prediction in gigapixel whole slide images," in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4015–4025, 2021.
- [14] N. Arya and S. Saha, "Multi-modal advanced deep learning architectures for breast cancer survival prediction," *Knowledge-Based Systems*, vol. 221, Art. no. 106965, 2021. doi: 10.1016/j.knosys.2021.106965.
- [15] S. Steyaert et al., "Multimodal data fusion for cancer biomarker discovery with deep learning," *Nature Machine Intelligence*, vol. 5, pp. 351–362, 2023. doi: 10.1038/s42256-023-00633-5.