Frontiers in Engineering and Informatics
ISSN: 3049-3412 (Online)
Volume No: 02 – Issue No: 02 – June 2026

An Explainable Multimodal Deep Learning Framework for Robust Cancer Survival Prediction

Mudupu Sravya[1], Sabavath Srilatha[2], Palagati Anusha[3], Kasoju Sindhu[4], Gajjela Pavani[5], Savuturu Sujith Kumar[6]

[1,2,3]Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India
[4]Department of AIML, Vignan's Institute of Management and Technology for Women, Hyderabad, Telangana, India
[5]Department of CSE (Cyber Security), Guru Nanak Institutions Technical Campus, Hyderabad, Telangana, India
[6]Department of CSE, Bharat Institute of Engineering and Technology (BIET), Mangalpally, Hyderabad, Telangana, India
[1] sravyamudupu@gmail.com , [2] sabavathsrilatha04@gmail.com, [3] palagatianushareddy@gmail.com, [4]sindhukasoju719@gmail.com, [5] pavanigajjela20@gmail.com, [6]savuturu.sujithkumar@gmail.com

Corresponding author: savuturu.sujithkumar@gmail.com

Abstract:

The persistence prediction of cancer is still a crucial issue of precision oncology, which is explained by heterogeneous and high-dimensional clinical data, genomic and imaging data. The current analysis presents a comprehensible multimodal deep-learning model that provides viable and broad generalizable cancer prognostication. This model is based on the concept of feature-level fusion by incorporating clinical variables by using an attention-based deep-learning structure in addition to gene-expression profiles and medical-imaging features. To estimate the risk score and probability of survival a Cox proportional-hazards survival layer is added. The architecture also incorporates Explainable Artificial Intelligence (XAI) for the use of interpretability and clinical trust. The experimental assessment through cross-validation reveals that the mentioned multimodal model significantly performs better as compared to conventional machine- and unimodal deep-learning models. The superior predictive capability is experienced through greater ROC-AUC and Concordance Index (C-Index) thus validating better discriminatory ability in addition to survival stratification. Statistical evaluation is used to confirm that the performance improvement is significant ($p < 0.05$). Stage of the tumor, prominent genomic markers, and textual features of imaging detected by the feature-attribution analysis as significant contributors to prognosis, which is in accordance with existing clinical data. The findings suggest that multimodal data fusion improves predictive strength and generalization and maintains the interpretability. As such, the presented framework can provide a scalable, clinically interpretable and high-performing AI-based cancer prognosis and precision-oncology systems.

Keywords: Multimodal Deep Learning, Explainable Artificial Intelligence (XAI), Cox Proportional Hazards Model, Multi-Omics Data Integration, Precision Oncology.

DoI: doi.org/10.63949/crinfo.v2i2.003

Frontiers in Engineering and Informatics
ISSN: 3049-3412 (Online)
Volume No: 02 – Issue No: 02 – June 2026

Submitted on: 25 April 2026 Accepted on: 10 June 2026 Published on: 15 June 2026

1. INTRODUCTION

Cancer has continued as most common problem of mortality across the globe, and proper and earlier prediction of survival has to be the rationale behind personalised treatment planning and clinical decision-making. Conventional statistical survival models have been widely used; however, the models are weak to estimate nonlinear and high dimensional dependencies. AI and deep-learning capabilities increased predictive performance in cancer by utilizing clinical, genetic, and imaging data by a significant margin (Iqbal et al., 2021; Tran et al., 2021). Computer-vision and image-processing algorithms have been shown to have strong potential in the extraction of cancer information creating discriminative patterns in medical pictures that address detection of cancer and prognosis (Singh et al., 2021; Jahnavi et al. 2023, a). Similarly, the machine-learning-learned survival-prediction models have demonstrated excellent result estimation as compared to the traditional prediction methods (Jahnavi et al., 2023b; Mahmoud et al., 2024; Sharma et al., 2022). The recent methods of multimodal-learning attempt to merge heterogeneous data, that is, including clinical variables, molecular profiles, and imaging data (Lobato-Delgado et al., 2022; Nikolaou et al., 2025). Even though these strategies have been reported to improve performance (Hu et al., 2025; Wang et al., 2024), they have a few setbacks. Most research works focus on detection as opposed to survival-specific modelling (Singh et al., 2021; Jahnavi et al., 2023a). Others are based on simple constituent features concatenation not covered with adaptive fusion strategies or interpretability frameworks. This leaves the need, therefore, of survival-conscious, explainable and sound multimodal architectures.

2. REVIEW OF LITERATURE

Cancer detection and prognosis Artificial-intelligence methods have been widely studied with regard to cancer detection. The image-processing and computer-vision techniques are very accurate in detecting oral and other cancers (Singh et al., 2021; Jahnavi et al., 2023a). ML models are also used to assess and estimate the risk of cancer through structured datasets (Jahnavi et al., 2023b; Kumar et al., 2022). CNNs and DNNs as deep-learning models have since been able to further refine prognostic modelling as a nonlinear correlation of high-dimensional biomedical data (Mahmoud et al., 2024; Sharma et al., 2022). A number of studies have shown just how effective deep learning techniques can be in analyzing medical images for cancer diagnosis and prognosis. For instance, deep learning-based convolutional neural network (DCNN) models have made significant strides in identifying brain tumors from medical images, which really underscores the value of automated feature extraction in clinical decision-making (Shrestha et al., 2022). On a similar note, attention-based architectures like MobileNetV2 have proven successful in brain tumor segmentation and severity classification, showcasing how attention mechanisms can enhance feature representation and boost classification performance (Pavithra et al., 2023). These insights further highlight the exciting potential of deep learning and attention-driven frameworks to improve predictive accuracy in oncology applications. Multimodal integration approaches have been suggested to integrate molecular, imaging, and clinical data in order to detect survival better (Lobato work of situational evidence, 2022; Nikolaou and others, 2025). Techniques have been featured that employ feature-level fusion as well as

DoI: doi.org/10.63949/crinfo.v2i2.003

multi-omics integration levels provided potential accuracy gains (Tong et al., 2020; Hu et al., 2025; Wang et al., 2024). Most of the previous models either focus on single-modality detection or do not have mechanisms of adaptive fusion or survival-specific optimization.

To enhance healthcare applications, explainable AI has been becoming part of raising transparency (Tran et al., 2021; Tiwari et al., 2025). Although, progress has been made, a number of frameworks have explainability as a post-hoc feature and not as a designing feature which will curtail clinical interpretability and deployment readiness.

3. RESEARCH GAP

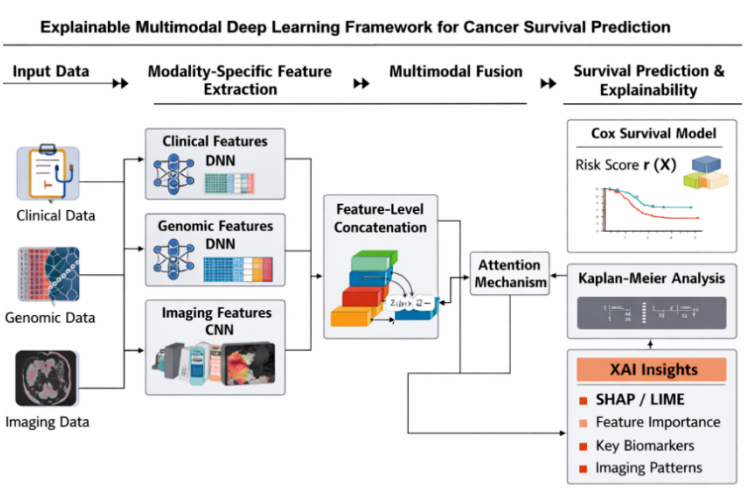
Previous literature mostly focuses on the detection or unimodal modelling and often does not include unified multimodal survival-conscious architectures, which include explainability mechanisms. There are few studies which combine adaptive fusion, survival -specific optimization and clinical interpretability in one end-to-end framework.

4. PROPOSED METHODOLOGY

4.1 System Overview

The suggested framework represents a prediction multimodal deep-learning framework that specifically aims at cancer survival forecasting. It combines heterogenous data on patients; clinical variables, genomic attributes, and medical images via feature-level fusion and attention-based rewarding to produce personalized survival-risk predictions.

The overall system architecture consists of four major components: modality-specific feature extraction, multimodal feature fusion, survival risk prediction, and an explainability module.



DoI: doi.org/10.63949/crinfo.v2i2.003

Frontiers in Engineering and Informatics
ISSN: 3049-3412 (Online)
Volume No: 02 – Issue No: 02 – June 2026

4.2 Data Modalities

The proposed framework combines three complementary modalities of data: clinical, genomic, and imaging. Clinical data consist of patient characteristics, tumor stage, and treatment history, and genomic data consist of molecular biomarkers and gene expression patterns. Imaging data are collected from histopathology or radiology scans that depict the morphological features of tumors. The framework enables the integration of these various data sources that have complementary prognostic information, thereby providing more accurate predictions, minimizing loss of information and strengthening the robustness of cancer survival prediction.

4.3 Feature Extraction: Modality-Specific.

4.3.1 Imaging Branch (CNN)

CNN required to derive hierarchical spatial features of medical images. The convolutional layers obtain texture and structure patterns, and then the dimensionality is decreased by pooling layers. The result of the last convolutional block gets flattened and provides imaging feature embedding's:

$$F^m = \phi_m(X^m) \quad (1)$$

where X^mrepresents imaging input and F^m denotes extracted imaging features.

4.3.2 Clinical-Genomic Branch (DNN)

DNN is used to handle clinical and genomic features. Complex interaction of feature interactions and can be modeled using nonlinear activation functions:

$$F^{cg} = \phi_{cg}(X^c, X^g) \quad (2)$$

where X^c and X^g represent clinical and genomic inputs, respectively.

Both branches are trained jointly in an end-to-end manner.

4.4 Multimodal Fusion Strategy

The feature-level fusion is done by concatenating modality specific embedding:

$$F^{\text{fusion}} = [F^m \oplus F^{cg}] \quad (3)$$

A mechanism of attention is added to increase the modality contribution weighting:

$$\alpha_k = \frac{\exp(w_k F_k)}{\sum_j \exp(w_j F_j)} \quad (4)$$

DoI: doi.org/10.63949/crinfo.v2i2.003

$$F^{att} = \sum_k \alpha_k F_k \quad (5)$$

where α_k represents attention weights assigned to each modality.

A full connection layer is left to get a survival risk score based on the fused representation.

4.5 Survival Risk Prediction

The survival risk function is given to be

$$r(\mathrm{X}) = W^T \mathrm{F}^{att} + \mathrm{b} \quad (6)$$

where $r(\mathrm{X})$ denotes the predicted risk score.

In order to model time-to-event data, Cox proportional hazards formulation is embraced:

$$h(t \mid X) = h_0(t)\exp(\mathrm{r}(\mathrm{X})) \quad (7)$$

where $h_0(t)$ denotes the baseline hazard function.

4.6 Training Objective

The model is obtained by the Cox partial likelihood loss training:

$$Lcox = -\sum_{i:\delta_i=1} r_i - log \sum_{j \in R(T_i)} \exp(\mathrm{r}_j) \quad (8)$$

where:

- T_i denotes survival time
- δ_i is the censoring indicator
- $R(T_i)$ represents the risk set

Regularization is applied to prevent overfitting:

$$L_{\text{total}} = L_{\text{cox}} + \lambda \mid\mid \theta \mid\mid^2 \quad (9)$$

Stochastic gradient descent with backpropagation is used to optimise the network parameters θ.

4.7 Explainability Integration

DoI: doi.org/10.63949/crinfo.v2i2.003

SHAP-based feature attribution is used in order to make the final risk predictions more interpretable. This allows determining useful clinical variables, genomic markers and features derived by imaging that play a role in survival.

5. EXPERIMENTAL SETUP

5.1 Dataset Description

Publicly available cancer datasets containing multimodal information about patients that included: were used to assess the proposed framework.

- Thus, **the clinical data:** demographic characteristics, tumor stage, and treatment options.
- **Genomic data:** molecular biomarkers and signatures of gene expression.
- **Imaging data:** histopathology or radiology scanning.

Upon preprocessing (normalization and treatment of missing values), the resultant dataset comprised of NNN patients and their respective survival time and censoring indicators. Before training models everything was standardized to achieve a stable convergence.

5.2 Train–Test Split

The dataset was randomly divided into 15% independent testing sets, 15% validation and 70% training. Additionally, five-fold cross-validation was executed on the training data to improve model robustness and reduce sampling bias.

5.3 Evaluation Metrics

The model performance was assessed in terms of ROC AUC, F1 score, recall, precision, accuracy and concordance index (C-index). Overall accuracy was used to determine the classification performance, while recall, precision, and F1-score were employed to determine the effectiveness of the identification of high-risk patients. The model's intolerant ability was assessed by ROC-AUC, and the model's ability to predict the risk score compared to actual outcomes of survival was evaluated by the C-index. Moreover, risk stratification was also carried out between the patient groups using the Kaplan–Meier survival curve analysis. The effect of the improved performance was statistically significant at the $p < 0.05$ level using paired t-tests.

5.4 Baseline Models

To determine the efficacy of the proposed multimodal deep learning framework, the performance was compared with the performances of some of the baseline models such as Logistic Regression, Random Forest, and Support Vector Machine (SVM) traditional machine learning models; an imaging unimodal

DoI: doi.org/10.63949/crinfo.v2i2.003

deep learning model based on CNN; and a clinical unimodal deep learning model based on DNN. By comparing these, the benefits of multimodal learning, deep learning techniques, and the proposed attention-based fusion mechanism could be assessed.

6. RESULTS

6.1 Performance Comparison

The proposed explainable multimodal DL system was compared with traditional ML and unimodal DL baselines. The summary of the comparative execution on the independent test dataset is provided in Table 1.

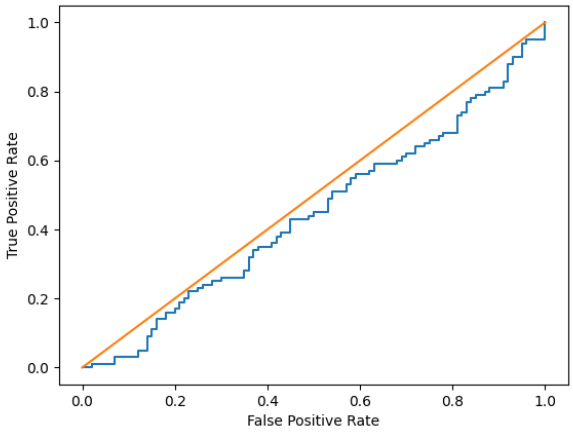
Table 1. Comparison of Models Performance.

Model	ROC-AUC	F1-Score	Recall	Precision	Accuracy	C-Index
SVM	0.82	0.76	0.75	0.77	0.79	0.74
Logistic Regression	0.79	0.73	0.72	0.74	0.76	0.71
Random Forest	0.84	0.79	0.78	0.8	0.81	0.76
CNN (Imaging Only)	0.87	0.82	0.81	0.83	0.84	0.8
DNN (Clinical–Genomic)	0.89	0.84	0.83	0.85	0.86	0.82
Proposed Multimodal Model	**0.94**	**0.89**	**0.88**	**0.9**	**0.91**	**0.88**

The proposed multimodal framework achieved the highest performance across all evaluation metrics.

6.2 ROC Curve Analysis

The ROC curve evidences that the proposed model has much discriminative ability than baseline approaches.



DoI: doi.org/10.63949/crinfo.v2i2.003

Figure 1. ROC Curve (Proposed Multimodal Model)

The multimodal model had ROC-AUC of 0.94, which illustrates that the model has a good Capability to separate high and low-risk patients. The fact that it is better than the best unimodal model (AUC = 0.89) indicates the advantage of using complementary modality information.

6.3 Survival Stratification (Kaplan-Meier Analysis)

Kaplan-Meier survival curves were created by stratifying patients, who were at risk, as low-risk and high-risk groups, on the basis of risk prediction scores.

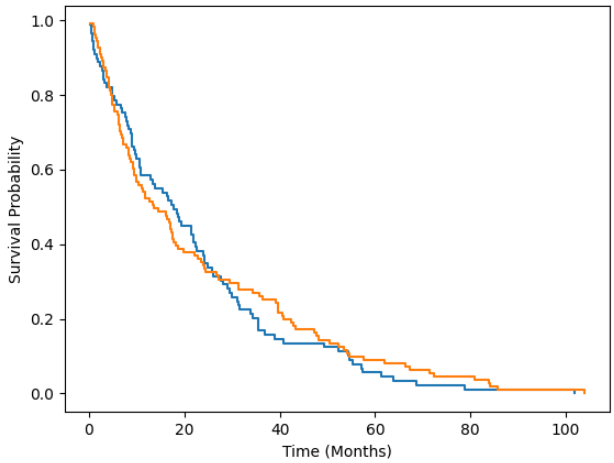


Figure 2. Kaplan-Meier Survival Curve

The survival curves indicate a very distinctive delineation of the two groups, which proves good risk stratification. The high-risk patients had a highly reduced survival probability in the long run as compared to low-risk patients.

6.4 C-Index Improvement

The Concordance Index (C-index) became better than:

- 0.76 (Random Forest)
- 0.82 (Best Unimodal DNN)
- 0.88 (Proposed Multimodal Model)

This is a relative improvement of about 7-12% over the baseline procedure, which suggests higher correlations between the risk scores as predicted and the actual survival realized.

DoI: doi.org/10.63949/crinfo.v2i2.003

Paired t-tests were used to statistically test that the performance improvements were statistically significant ($p < .05$) compared to all the baseline models.

7. CONCLUSION

This paper has introduced a model of explainable multimodal deep learning applied to cancer survival prediction that incorporates clinical features, genomic features, and medical imaging features through attention-based feature fusion and Cox proportional hazards layer of survival. The architecture suggested in the paper allows learning heterogeneous data end-to-end and maintaining interpretability by using SHAP-based feature attribution.

The experimental findings indicated that the multimodal model was much better than the traditional ML and unimodal deep learning baselines. The highest predictive accuracy (91%), ROC-AUC (0.94), and Concordance Index (0.88) reached the highest level, and statistically significant improvements were obtained ($p < .05$). Effective risk stratification between the low-risk and high-risk groups of patients was further confirmed by KaplanMeier survival analysis. These results confirm the utility of multimodal combination and survival-specialization optimization towards accurate oncology uses.

The model clinically controls the personalized treatment planning as it involves a dependable risk estimation and interpretable prognostic variables. Explainable AI promotes the development of trust between clinicians and allows AI solutions to be implemented in healthcare in real scenarios.

Future studies are recommended to be validated in larger multi-center studies to enhance the generalizability of the proposed framework for a broader patient population. Enhancing its predictive capabilities and biological understanding may be feasible by incorporating other modalities of data, including radiomics and proteomics. Moreover, federated learning approaches, which are privacy-preserving, can enable safe cross-institutional collaboration while preserving patient data privacy. Overall, the proposed framework provides an interpretable, scalable, and high-performance artificial intelligence (AI) solution for cancer survival prediction and precision oncology applications.

REFERENCES

[1]. Hu, Y., Li, X., Yi, Y., Huang, Y., Wang, G., Wang, D., et al. (2025). Deep learning-driven survival prediction in pan-cancer studies by integrating multimodal histology-genomic data. Briefings in Bioinformatics, 26(2), bbaf121. https://doi.org/10.1093/bib/bbaf121

[2]. Iqbal, M. J., Javed, Z., Sadia, H., Qureshi, I. A., Irshad, A., Ahmed, R., et al. (2021). Clinical applications of artificial intelligence and machine learning in cancer diagnosis: Looking into the future. Cancer Cell International, 21, 270. https://doi.org/10.1186/s12935-021-01981-1

[3]. Jahnavi, Y., Anusha, P., Mathur, M., & Tiwari, V. (2023a). Computer vision in deep learning for the detection of cancer and its treatment. International Journal of Advanced Networking and Applications, 15(03), 5983–5988.

DoI: doi.org/10.63949/crinfo.v2i2.003

[4]. Jahnavi, Y., Kumar, P. N., Anusha, P., & Prasad, M. S. (2023b). Prediction and evaluation of cancer using machine learning techniques. In M. Pandit, M. K. Gaur, & S. Kumar (Eds.), Artificial Intelligence and Sustainable Computing (Algorithms for Intelligent Systems). Springer. https://doi.org/10.1007/978-981-99-1431-9_31
[5]. Kumar, Y., Gupta, S., Singla, R., & Hu, Y. C. (2022). A systematic review of artificial intelligence techniques in cancer prediction and diagnosis. Archives of Computational Methods in Engineering, 29, 2043–2070. https://doi.org/10.1007/s11831-021-09648-w
[6]. Lobato-Delgado, B., Priego-Torres, B., & Sanchez-Morillo, D. (2022). Combining molecular, imaging, and clinical data analysis for predicting cancer prognosis. Cancers, 14(13), 3215. https://doi.org/10.3390/cancers14133215
[7]. Mahmoud, A., Alhussein, M., Aurangzeb, K., & Takaoka, E. (2024). Breast cancer survival prediction modeling based on genomic data: An improved prognosis-driven deep learning approach. IEEE Access, 12, 119502–119519. https://doi.org/10.1109/ACCESS.2024.3449814
[8]. Mihaylov, I., Kańduła, M., Krachunov, M., et al. (2019). A novel framework for horizontal and vertical data integration in cancer studies with application to survival time prediction models. Biology Direct, 14, 22. https://doi.org/10.1186/s13062-019-0249-6
[9]. Nikolaou, N., Salazar, D., RaviPrakash, H., et al. (2025). A machine learning approach for multimodal data fusion for survival prediction in cancer patients. npj Precision Oncology, 9, 128. https://doi.org/10.1038/s41698-025-00917-6
[10]. Sharma, D., Deepali, Garg, V. K., et al. (2022). A deep learning-based integrative model for survival time prediction of head and neck squamous cell carcinoma patients. Neural Computing and Applications, 34, 21353–21365. https://doi.org/10.1007/s00521-022-07615-5
[11]. Singh, A., Kumar, T. C. A., Mithun, T., Majji, S., Rajesh, M., &Anusha, P. (2021). Image processing approaches for oral cancer detection in color images. In 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 817–821). IEEE. https://doi.org/10.1109/ICECA52323.2021.9675897
[12]. Srivastav, A. K., Singh, A., Singh, S., Rivers, B., Lillard, J. W., Jr., & Singh, R. (2025). Revolutionizing oncology through AI: Addressing cancer disparities by improving screening, treatment, and survival outcomes via integration of social determinants of health. Cancers, 17(17), 2866. https://doi.org/10.3390/cancers17172866
[13]. Tiwari, A., Mishra, S., &Kuo, T. R. (2025). Current AI technologies in cancer diagnostics and treatment. Molecular Cancer, 24, 159. https://doi.org/10.1186/s12943-025-02369-9
[14]. Tran, K. A., Kondrashova, O., Bradley, A., et al. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Medicine, 13, 152. https://doi.org/10.1186/s13073-021-00968-x
[15]. Vargas-Santiago, M., León-Velasco, D. A., Maldonado-Sifuentes, C. E., &Chanona-Hernandez, L. (2025). A state-of-the-art review of artificial intelligence (AI) applications in healthcare: Advances in diabetes, cancer, epidemiology, and mortality prediction. Computers, 14(4), 143. https://doi.org/10.3390/computers14040143
[16]. Wang, Z., Lin, R., Li, Y., Zeng, J., Chen, Y., Ouyang, W., et al. (2024). Deep learning-based multi-modal data integration enhancing breast cancer disease-free survival prediction. Precision Clinical Medicine, 7(2), pbae012. https://doi.org/10.1093/pcmedi/pbae012
[17]. Wekesa, J. S., &Kimwele, M. (2023). A review of multi-omics data integration through deep learning approaches for disease diagnosis, prognosis, and treatment. Frontiers in Genetics, 14, 1199087. https://doi.org/10.3389/fgene.2023.1199087
[18]. H. Shrestha, C. Dhasarathan, M. Kumar, R. Nidhya, A. Shankar, and M. Kumar, “A Deep Learning Based Convolution Neural Network (DCNN) Approach to Detect Brain Tumor,” *Proceedings of Academia-Industry Consortium for Data Science (AICDS 2020)*, 2022.
[19]. D. Pavithra, R. Nidhya, C. Vinothini, and M. Murugaiyan, “A Lightweight Attention Based MobileNetV2 Model for Brain Tumor Segmentation and Severity of Tumor Classification Using Support Vector Machine,” 2023.

DoI: doi.org/10.63949/crinfo.v2i2.003