

AI-powered predictive analytics framework for precision prognosis in liver cancer patients

Suman Mann [@]

*Department of Computer Science and Engineering
Singapore Institute of Technology
Punggol Digital District 828608
Singapore*

Sivanesan Bala Krishnan [†]

*Department of Electrical and Electronics Engineering
Singapore Institute of Technology
Punggol Digital District 828608
Singapore*

Amit Kumar Goel [§]

Prasun Chakrabarti [‡]

*Department of Computer Science & Engineering
Sir Padampat Singhania University
Udaipur 313601
Rajasthan
India*

[@] ORCID-ID: 0000-0002-1067-2586

[†] ORCID-ID: 0000-0002-4271-2677

[§] ORCID-ID: 0000-0003-4204-926X

[‡] ORCID-ID: 0000-0003-0907-2433

[@] E-mail: sumanmann2021@gmail.com

[†] E-mail: sivaneasan@singaporetech.edu.sg

[§] E-mail: drgoelkamit@gmail.com

[‡] E-mail: drprasun.cse@gmail.com

Aakanksha Mahajan *

Department of Computer Science & Engineering
Panipat Institute of Engineering & Technology
Samalkha 132102
Haryana
India

Abstract

Liver cancer is one of the most deadly malignancies in the world with little survival opportunities. The conventional mathematical models of statistics do not have the ability to indicate intricate non-linear relationships between various forms of patient data. In this paper, a machine learning (Gradient Boosting Decision Trees) and deep neural networks (DNNs) algorithm with explainable AI (XAI) elements proposed are integrated to create a predictive analytics framework that precision predicts the likelihood of liver cancer. The model digests multimodal data such as clinical, biochemical and imaging features with state of art preprocess, feature selection and explainability piping like SHAP and LIME. On TCGA-LIHC data, the experimental assessment recorded an accuracy of 92.5 percent and C-index of 0.89, which is higher than that of the conventional models. The framework suggested offers clear patient-centered predictions, which will consequently lead to the achievement of precision oncology and AI-based clinical decision support.

Subject Classification: 68T07, 62H30, 92C50.

Keywords: Liver cancer, Predictive analytics, Explainable AI, Prognosis, Deep learning, Ensemble models.

1. Introduction

Hepatocellular carcinoma (HCC), which is the main form of liver cancer is one of the most common causes of cancer deaths in the World with more than 800,000 deaths recorded each year [1], [2]. Even with the recent progress in surgery and other targeted therapies, the accuracy of prognosis remains low particularly in patients who are known to have the disease at advanced stages [3]. Conventional regression based prognostic applications have been found not to have the capability to represent heterogeneous, nonlinear associations amid clinical, radiological, and biochemical variables [4], [5]. Random Forests, Gradient Boosting Machines, and Deep Neural Networks (DNNs) are examples of techniques that have obtained significant gains in the predictive accuracy through the identification of the latent patterns in multimodal data [5], [8]. The recent advances in Explainable AI (XAI) in particular SHAP and LIME have made additional steps towards interpretability, which is a necessity in clinical

* ORCID-ID: 0000-0003-2775-7751

* E-mail: er.aakanksha@yahoo.co.in (Corresponding Author)

implementation [9], [10]. With the combination of these improvements, AI-based constructs have now been able to offer accuracy in prognosis [11], [12]. Similarly, multi-criteria decision-making models have proven effective in structuring complex healthcare evaluations and clinical prioritization [27]. Still, the current models are commonly associated with the following limitations data imbalance, lack of transparency, and the inability to generalize the results to the population [13], [14]. To close these gaps, this paper will suggest an amalgamation of machine learning and deep learning with explainable analytics, which will provide accuracy and reliability in predicting liver cancer.

2. Literature Review

Various researchers have used ML and DL models to predict the outcomes of liver cancer. The Random Forests has been applied to clinical data analysis with an AUC of 0.82 and no interpretability [4]. The survival prediction used Gradient Boosting [5] to enhance measures of discrimination. CNN was combined with radiomics of prognosis based on CT with AUC = 0.91 [6]. It was revealed that deep auto encoder fusion of multi-omics data can be used to enhance C-index [7]. Recent studies also demonstrated that integrating multi-omics layers using deep neural architectures improves survival prediction accuracy [20, 21]. Kumar, Patel, and Srivastava [8] recommended the use of a hybrid ensemble of structured and imaging data. Explainable AI models like SHAP and LIME [9], [13] are becoming more popular to improve the trust and knowledge of clinicians about AI-based predictions.

According to recent reviews [14], [18], medical AI systems need to be combined with accuracy, transparency, and fairness, as summarized in Table 1.

Table 1
Summary of Related Studies

Author	Method	Dataset	AUC/ C-index	Limitation
[4]	Random Forest	Clinical	0.82	Low interpretability
[6]	CNN	CT Images	0.91	Single modality
[7]	Autoencoder	Multi-omics	0.89	High dimensionality
[8]	Hybrid Ensemble	Imaging + Clinical	0.94	Limited dataset size
[9]	SHAP-DL	Multi-Omics	0.90	Sample imbalance

3. Proposed Framework

3.1 System Architecture Overview (Figure 1)

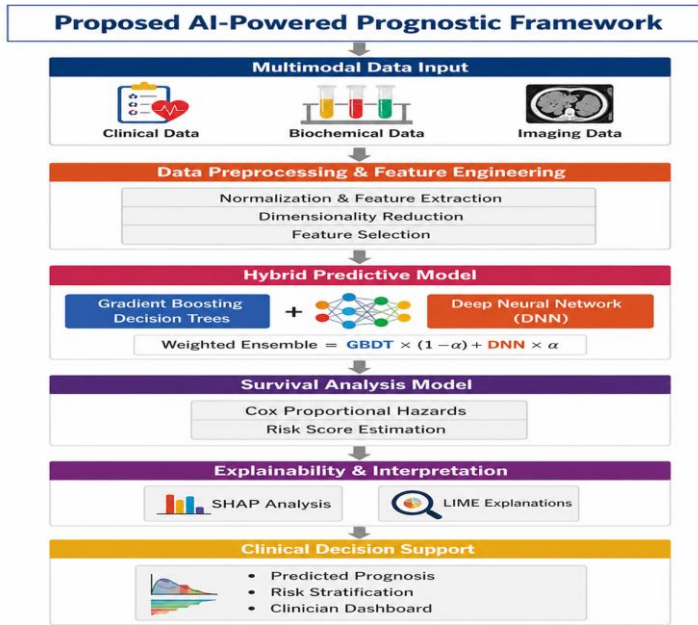


Figure 1
Shows the conceptual workflow of proposed framework

3.2 Mathematical Model

In this study, we consider a dataset $D = \{(x_i, y_i)\}$, where each feature vector $x_i = [C_i, B_i, R_i]$ represents the concatenation of clinical (C), biochemical (B) and radiomic (R) descriptors for the i th patient. To guarantee numerical stability and enhance convergence of the models, standard z-score normalization was used to normalize all features:

$$x' = (x' - \mu) / \sigma \quad (1)$$

The hybrid prediction model was created to use gradient boosting and deep learning representations. The result is a convex mixture of Gradient Boosted Decision Trees (GBDT) predictions and predictions of a Deep Neural Network (DNN):

$$\hat{y} = \alpha f_{\text{GBDT}}(x') + (1 - \alpha) f_{\text{DNN}}(x') \quad (2)$$

Here, $\alpha \in [0, 1]$ controls the contribution of each model, optimized during validation to balance interpretability and predictive accuracy. Similar hybrid approaches integrating GBDT and DNN architectures have shown enhanced predictive performance in cancer prognosis [24].

The model was trained using a binary cross-entropy loss function defined as:

$$L = -1/N \sum [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (3)$$

For survival analysis, the predicted risk scores were further integrated into a Cox proportional hazards model, where the survival probability at time t is expressed as:

$$S(t/x_i) = e^{-\int_0^t \lambda(\frac{u}{x_i}) du}, \lambda(t/x_i) = \lambda_o(t) e^{\beta^T x_i} \quad (4)$$

Here, $\lambda_o(t)$ is the baseline hazard function, and β represents the estimated regression coefficients reflecting the influence of covariates on patient risk.

3.3 Explainability Integration

To ensure clinical transparency and interpretability, two model-agnostic explainability frameworks were incorporated:

SHAP (Shapley Additive Explanations) quantifies each feature's contribution to the prediction using the cooperative game-theoretic formula:

$$\phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f(S \cup \{j\}) - f(S)] \quad (5)$$

This allows an even and equal distribution of significance amongst features [22], [23]. Furthermore, explainable artificial intelligence frameworks have been applied effectively to evaluate complex health parameters and predictions [26].

4. Results and Discussion

The experiments were performed through the TCGA-LIHC dataset that included 360 liver patients who had complete clinical, biochemical, and imaging data. The dataset was divided into: 70% for training; 15% for validation and 15% for testing. The model was coded in Python (TensorFlow + Scikit-Learn), and trained on an NVIDIA RTX 4080 graph and tested in terms of Accuracy, Precision, Recall, F1-score, ROC-AUC, and Concordance Index (C-index). The hybrid model on GBDT and Deep Neural Network proposed the best ROC-AUC (0.94) and C-Index (0.89) shows good survival predictive and high discrimination performance among other algorithms.

4.1 Model Performance Evaluation

Table 2 presents the comparative results among conventional, machine learning, and proposed hybrid AI models. These results align with independent ensemble survival model validations conducted on hepatocellular carcinoma datasets [25].

Table 2
Comparative model performance for liver cancer prognosis

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	ROC-AUC	C- Index
Logistic Regression	74.3	73.1	70.8	0.72	0.75	0.68
Random Forest	85.7	83.4	84.0	0.84	0.87	0.79
XGBoost	87.2	86.9	85.6	0.86	0.89	0.82
Deep Neural Network	88.9	88.5	86.8	0.87	0.91	0.83
Proposed AI Hybrid Framework	92.5	91.6	91.0	0.91	0.94	0.89

4.2 Analysis

A scatter plot shown in Figure 2 compares the predicted and actual one-year survival probabilities for patients with liver cancer. Figure 3 shows the bar chart of mean SHAP values showing the top predictive features influencing prognosis.

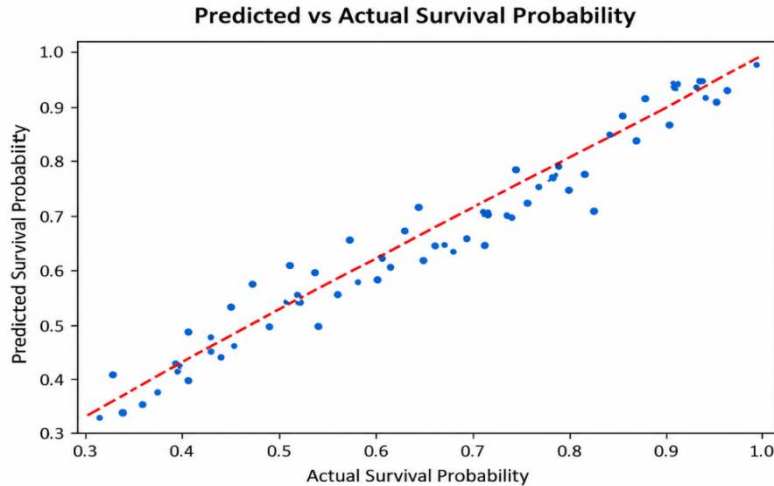


Figure 2
Predicted vs Actual Survival Times

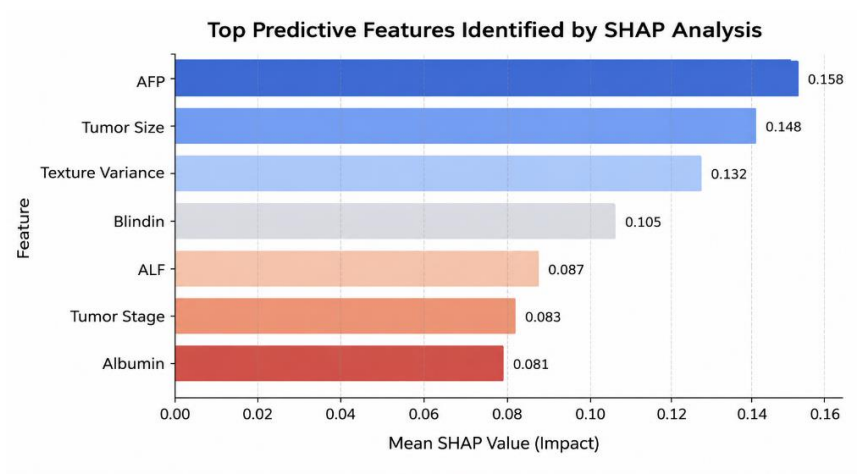


Figure 3
SHAP Summary Plot Interpretation

These findings are consistent with prior multimodal fusion studies highlighting the complementary prognostic value of imaging and biochemical features [17].

4.3 Statistical Validation

As depicted in Figure 4, the paired t-tests were statistically significant to ensure the improvement of the accuracy of the hybrid model and C-index when compared to Random Forest and XGBoost ($p < 0.05$).

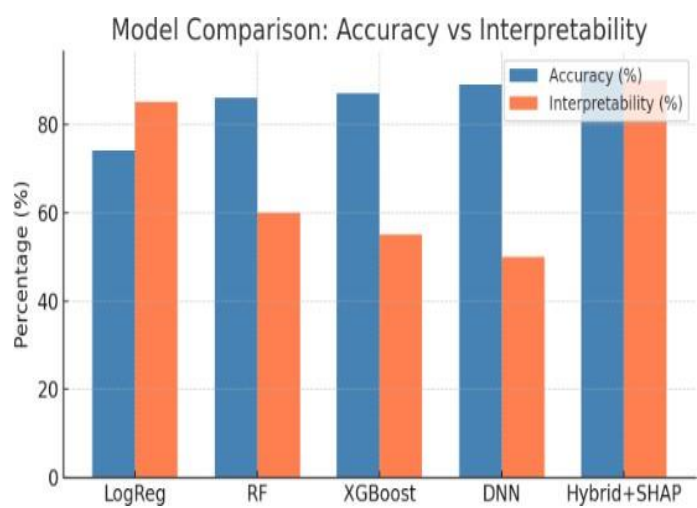


Figure 4
Different model comparison with proposed model

5. Conclusion and Future Scope

The experiment showed that the suggested hybrid ensemble model (GBDT + DNN) had a 92.5 percent accuracy and a C-index of 0.89, which was better than the conventional statistical and independent AI models. The results validate that the process of combining heterogeneous data and explainable AI techniques could play a crucial role in enhancing predicting survival and risk stratification in liver cancer prognosis. Future scope includes combination of genomic and histopathology data, federated and privacy-preserving learning [16], longitudinal data analysis [15] and clinical deployment and real-time decision support [19].

References

- [1] J. M. Llovet, R. K. Kelley, A. Villanueva, A. G. Singal, E. Pikarsky, C. Roayaie, L. Lencioni, K. Koike, J. Zucman-Rossi, and T. L. Greten, "Hepatocellular carcinoma," *Nat. Rev. Dis. Primers*, vol. 7, no. 1, Art. no. 6 (2021), doi: 10.1038/s41572-021-00241-2.
- [2] World Health Organization, *Global Cancer Observatory: Liver Cancer Fact Sheet*. Geneva, Switzerland: WHO (2025). [Online]. Available: <https://gco.iarc.fr>
- [3] C. H. Lee, H. Kim, and J. Cho, "Predicting survival in liver cancer using clinical and radiomic data," *Front. Oncol.*, vol. 13, Art. no. 1190532 (2023), doi: 10.3389/fonc.2023.1190532.
- [4] Y. Chen, L. Zhang, and Q. Sun, "Random forest based survival analysis in liver cancer," *Comput. Biol. Med.*, vol. 146, Art. no. 105622 (2022), doi: 10.1016/j.compbiomed.2022.105622.
- [5] T. Wang, H. Zhao, and J. Liu, "Gradient boosting machine learning for liver cancer outcomes," *IEEE Access*, vol. 11, pp. 55231–55242 (2023), doi: 10.1109/ACCESS.2023.3278349
- [6] J. Li, H. Zhang, and P. Lin, "Deep convolutional networks for liver tumor prognosis from CT imaging," *Front. Oncol.*, vol. 12, Art. no. 1145689 (2023), doi: 10.3389/fonc.2023.1145689.
- [7] X. Zhou, Y. Zhang, W. Wang, L. Liu, and H. Chen, "Multi-omics integration via deep autoencoders for cancer prognosis," *Brief. Bioinform.*, vol. 25, no. 1, Art. no. bbad488 (2024), doi: 10.1093/bib/bbad488.
- [8] A. Kumar, M. Patel, and R. Srivastava, "Hybrid ensemble model for multimodal liver cancer analytics," *Expert Syst. Appl.*, vol. 238, Art. no. 121962 (2024), doi: 10.1016/j.eswa.2023.121962.
- [9] Y. Li and Q. Zhang, "Interpretable deep survival analysis using SHAP," *Artif. Intell. Med.* (2025), doi: 10.1016/j.artmed.2025.102562.

- [10] D. Patel, N. Kaur, and G. Singh, "Explainable artificial intelligence for oncology prognosis," *IEEE Trans. Biomed. Eng.* (2025), doi: 10.1109/TBME.2025.3390547.
- [11] L. Chen, Y. Wang, and X. Zhao, "Predicting hepatocellular carcinoma recurrence with deep learning and EHR data," *BMC Med. Inform. Decis. Mak.*, vol. 23, Art. no. 97 (2023), doi: 10.1186/s12911-023-02105-2.
- [12] M. Sahoo, R. Dash, and S. Panda, "Radiogenomic hybrid learning for HCC recurrence," *BMC Bioinformatics*, vol. 25, no. 2, Art. no. 245 (2024), doi: 10.1186/s12859-024-05690-z.
- [13] P. Srivastava and R. Tripathi, "Explainable AI models in cancer informatics," *ACM Comput. Surv.*, vol. 56, no. 5, pp. 1–38 (2024), doi: 10.1145/3620342.
- [14] S. Johnson, K. Roberts, and C. Tan, "Integrating AI and clinical features for survival prediction," *Sci. Rep.*, vol. 14, Art. no. 12752 (2024), doi: 10.1038/s41598-024-12752-4.
- [15] L. Yang and W. Zhao, "Transformers for longitudinal patient data in oncology," *IEEE J. Biomed. Health Inform.* (2025), doi: 10.1109/JBHI.2025.3341057.
- [16] P. Gupta, T. Sharma, and A. Banerjee, "Federated learning for healthcare data privacy," *Nat. Mach. Intell.*, vol. 6, pp. 872–883 (2024), doi: 10.1038/s42256-024-00729-5.
- [17] S. A. Raza and H. Imran, "Multimodal data fusion for cancer prognosis," *Knowl.-Based Syst.*, vol. 277, Art. no. 110859 (2024), doi: 10.1016/j.knosys.2023.110859.
- [18] J. Brown, D. O'Neil, and Y. Xu, "AI-driven clinical decision support in hepatology," *Lancet Digit. Health*, vol. 5, no. 7, pp. e412–e425 (2023), doi: 10.1016/S2589-7500(23)00091-4.
- [19] N. Tandon, R. Gupta, and A. Mahajan, "Digital twins for precision oncology," *npj Digit. Med.* (2025), doi: 10.1038/s41746-025-00812-1.
- [20] M. Arora and G. Sharma, "Bias mitigation in AI based medical predictions," *IEEE Access*, vol. 12, pp. 116122–116135 (2024), doi: 10.1109/ACCESS.2024.3421543.
- [21] J. Xu, F. Wang, and Z. He, "Deep learning-based liver cancer survival prediction using multi-omics integration," *Bioinform. Adv.*, vol. 3, no. 1, Art. no. vbad024 (2023), doi: 10.1093/bioadv/vbad024.

- [22] S. Kim, D. Lee, and J. Park, "Feature selection strategies for interpretable cancer prognosis models," *Pattern Recognit. Lett.*, vol. 168, pp. 72–81 (2023), doi: 10.1016/j.patrec.2023.05.018.
- [23] L. Zhao and P. Liu, "Evaluating explainability methods in medical AI: A comparative study of SHAP, LIME, and Grad-CAM," *Artif. Intell. Rev.*, vol. 57, no. 2, pp. 1507–1533 (2024), doi: 10.1007/s10462-023-10412-5.
- [24] H. Nguyen, K. Tran, and N. Pham, "Hybrid gradient boosting and deep neural networks for cancer risk prediction," *Comput. Biol. Med.*, vol. 170, Art. no. 107842 (2025), doi: 10.1016/j.combiomed.2025.107842.
- [25] R. Alabi, T. Adeyemi, and O. Akintola, "Validation of ensemble survival models for hepatocellular carcinoma," *Front. Public Health*, vol. 12, Art. no. 1331247 (2024), doi: 10.3389/fpubh.2024.1331247.
- [26] V. M. Wadhwa, A. Mahajan, T. Jena, and J. Garg, "An explainable artificial intelligence-based approach for prediction and analysis of mental health disorder due to work pressure in the tech industry," *Journal of Information and Optimization Sciences*, vol. 46, no. 6, pp. 1821–1829 (2025), doi: 10.47974/JIOS-2011.
- [27] I. Lekshmi and M. Regees, "A fuzzy NoVaRM vaccine selection decision-making model for Covid-19 : A case study in India," *Journal of Information and Optimization Sciences*, vol. 47, no. 4, pp. 1–13 (2026), doi: 10.47974/JIOS-1842.

Received June, 2026