Predicting Breast Cancer Survival: An Approach using Deep Learning and Machine Learning Techniques

Soma Sekhar Kolisetty, Parlapalli Haseena, Chennupalli Chandrika Tirumala, Kodavati Jayamma,
Kathi Chandra Mouli, Anitha Vulugundam, K.V.Narasimha Reddy
Department of Computer Science and Engineering,
Narasaraopeta Engineering College (Autonomous)
Yellamanda Road, Narasaraopet – 522601, Andhra Pradesh, India
sekhar.soma007@gmail.com, haseenaparlapalli17@gmail.com, tirumalachennupalli@gmail.com, k.jayamma879@gmail.com,
chandramouli1714@grietcollege.com, anitha.vulugundam@gnits.ac.in, narasimhareddynec03@gmail.com

***Abstract*—Accurate prediction of breast cancer survival is a critical task for guiding treatment planning and clinical decision-making. In this study, we present a comparative analysis of classical survival models, machine learning techniques, and deep learning architectures using the METABRIC dataset. Traditional approaches, including the Cox Proportional Hazards (CPH) model and Random Forest Survival (RFS), achieved the highest predictive performance, with C-indices of 0.8719 and 0.8249, respectively. We further introduce SA-DGNet, a novel hybrid deep learning framework that integrates gated neural networks with self-attention mechanisms to capture both short-term and long-range temporal dependencies. While SA-DGNet obtained a lower C-index (0.7590) compared to CPH and RFS, it demonstrated enhanced interpretability, personalized risk stratification, and temporal attention visualizations that are clinically meaningful. Our findings highlight that classical models remain highly competitive for structured datasets, but deep learning models such as SA-DGNet open new avenues for multimodal integration, time-varying patient trajectories, and explainable prognosis tools. This work contributes a systematic benchmark across modeling paradigms and lays the groundwork for future research in interpretable deep survival prediction.**

***Index Terms*—Index Terms— Survival Analysis, SA-DGNet, Self-Attention, Gated Neural Network, Breast Cancer Prognosis, Deep Learning**

I. Introduction

Breast cancer remains one of the leading causes of mortality among women worldwide, and accurate survival prediction plays a vital role in supporting clinical decision-making, treatment planning, and patient counseling. Survival analysis, also referred to as time-to-event analysis, is particularly well suited for this task as it not only estimates the likelihood of an outcome but also provides insight into the timing of such events. The ability to model censored data, where outcomes are partially observed, further enhances its importance in medical prognosis.

Traditional statistical models, such as the **Cox Proportional Hazards (CPH)** model, have long served as benchmarks due to their interpretability and robustness in structured clinical datasets. More recently, machine learning methods like **Random Forest Survival (RFS)** have demonstrated strong performance by capturing nonlinear feature interactions and offering improved risk stratification. In parallel, the rise of **deep learning** has led to architectures such as **DeepSurv** and **DeepHit**, which extend survival analysis through neural networks capable of modeling complex patterns. However, these models often require large-scale or multimodal data to surpass classical approaches.

Despite these advances, two challenges remain open. First, classical models often underutilize temporal patterns and high-dimensional genomic data, potentially limiting their applicability in dynamic patient monitoring. Second, while deep learning approaches promise flexibility, their performance is often hampered by overfitting and lack of interpretability in medium-sized clinical datasets such as METABRIC. This motivates the need for novel hybrid frameworks that balance predictive accuracy, interpretability, and temporal modeling.

To address this gap, we introduce **SA-DGNet (Self-Attentive Deep Gated Network)**, a deep survival model that integrates **gated neural layers** to regulate information flow and **self-attention mechanisms** to capture long-range dependencies. SA-DGNet is designed not only to predict survival outcomes but also to provide interpretable attention visualizations that highlight critical time points and features influencing patient risk.

Contributions of this paper

The main contributions of this work can be summarized as follows:

- **Comprehensive Benchmarking:** We systematically evaluate statistical (CPH), machine learning (RFS), and deep learning (DeepSurv, DeepHit, SA-DGNet) approaches on the METABRIC breast cancer dataset.
- **Proposed Model:** We present SA-DGNet, a hybrid gated-attention network that learns both short-term and long-range dependencies, enhancing interpretability in survival prediction.
- **Empirical Findings:** Our experiments demonstrate that classical models such as CPH (C-index: 0.8719) and RFS

(0.8249) outperform current deep learning models, but SA-DGNet offers meaningful interpretability and personalization for patient risk stratification.
- **Future Potential:** We outline directions where SA-DGNet can be extended, including multimodal fusion of genomic, imaging, and clinical data, time-varying health trajectories, and real-time clinical decision support systems.

This paper is organized as follows: Section II reviews related work, Section III describes the methodology and proposed model, Section IV details the algorithm and experimental setup, Section V presents results and analysis, and Section VI concludes with key findings and future research opportunities.

The survival function is defined to characterize variability in event occurrence over time :

$$S(t|x) = \mathbb{P}(T > t \mid x) \tag{1}$$

It captures the variability in survival probabilities among individuals with heterogeneous covariates x beyond time t. This function is particularly valuable in medical applications, as it provides a time-varying estimate of the likelihood of survival over the course of heterogeneous follow-up durations among patients.

The hazard function characterizes the variability in instantaneous event rates across individuals, representing the probability that the event occurs at time t given it has not occurred before t:

$$h(t|x) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T < t + \Delta t \mid T \geq t, x)}{\Delta t} \tag{2}$$

To capture individual-specific risk patterns, the Cox Proportional Hazards (CPH) model , a semi-parametric model, is widely employed to represent heterogeneous hazard functions as:

$$h(t|x) = h_0(t) \cdot \exp(x^\top \beta) \tag{3}$$

Here, $h_0(t)$ is the baseline hazard function, and β is a learned coefficient vector. This model assumes that the effect of covariates on the hazard is log-linear and remains constant over time (proportional hazards). Despite these assumptions, CPH [4]remains popular due to its interpretability, analytical clarity, and strong performance in clinical datasets with moderate complexity of characteristics.cteristics.

To handle non-linear interactions and better adapt to real-world data, we implemented Random Forest Survival (RFS) [10], a non-parametric ensemble method based on decision trees. RFS constructs decision trees on heterogeneous bootstrapped samples to model dissimilar data structures, and combines their outputs to estimate the cumulative hazard function or survival probabilities. It does not rely on any parametric assumptions and is inherently robust to missing data and outliers. Moreover, RFS supports built-in [6]feature importance estimation, offering a degree of interpretability.

In our study, we evaluated both CPH and RFS [11] on the METABRIC dataset—a well-known breast cancer cohort with over 1,900 samples containing both clinical and genomic features. Our experiments revealed that **CPH achieved the highest Concordance Index (C-index)** among all models tested, followed closely by RFS. This demonstrates that even with the emergence of deep learning, classical statistical and tree-based methods continue to dominate in structured biomedical datasets.

However, real-world patient data often contain longitudinal patterns, nonlinear trends, and temporally varying covariates that are not fully captured by CPH or RFS [8]. To address this gap, recent advancements have introduced and deep learning-based models such as DeepSurv and DeepHit, which address dissimilar survival dynamics.DeepSurv substitutes the linear term in the Cox model with a neural network to capture personalized, non-linear risk functions, enabling the modeling of nonlinear relationships. DeepHit [12], on the other hand, directly estimates the joint distribution of survival time using a discrete-time approach. Although these models are theoretically more expressive, they did not outperform CPH or RFS in our METABRIC [13] data set, probably due to overfitting or insufficient temporal granularity.

To further explore the potential of deep learning in survival analysis, we propose **SA-DGNet (Self-Attentive Deep Gated Network)**—a novel architecture specifically designed to model complex temporal dependencies in patient health trajectories. SA-DGNet addresses key limitations in prior work by incorporating:

- **Gated Layers**, inspired by highway networks and LSTMs, to dynamically regulate information flow and highlight relevant temporal features at each stage of the sequence.
- **Self-Attention Mechanisms**, adapted from transformer models, to model non-local dependencies and provide interpretability by identifying which time steps contribute most to the prediction.

This dual mechanism enables SA-DGNet to compute time-adaptive risk scores, generate personalized survival curves, and produce attention visualizations that help clinicians understand the reasoning behind predictions.

Despite not outperforming CPH and RFS in terms of C-index, SA-DGNet demonstrated valuable capabilities in:

- Modeling patient-specific time series data,
- Learning temporal patterns in high-dimensional clinical features, and
- Providing interpretable risk stratification insights over time.

These characteristics make SA-DGNet a strong foundation for future survival models aimed at multimodal data integration, treatment response forecasting, or real-time prognosis in clinical decision support systems.

II. RELATED WORK

A. Statistical Survival Models

The Cox Proportional Hazards (CPH) model is a widely used semi-parametric [15] method in survival analysis. It

models the hazard function as:

$$h(t|x) = h_0(t) \cdot \exp(x^\top \beta) \tag{4}$$

where $h_0(t)$ is the baseline hazard function and β is the coefficient vector. Despite its linearity and proportional hazards assumption, CPH remains effective in structured datasets. In our METABRIC-based study [16], CPH achieved the highest concordance index (C-index), showing strong predictive ability in clean clinical data.

B. Machine Learning for Survival Prediction

Random Forest Survival (RFS) is a tree-based ensemble method that models cumulative hazard functions without assuming proportionality. RFS handles nonlinearities and missing values naturally. In our experiments, RFS performed comparably to CPH [17], confirming its robustness and adaptability to high-dimensional clinical features.

C. Deep Learning in Survival Analysis

DeepSurv and DeepHit [18] are neural network-based models designed to overcome CPH limitations. DeepSurv replaces the linear risk function with a nonlinear neural network, while DeepHit estimates discrete-time survival distributions. Despite their theoretical strengths, both models underperformed compared to CPH and RFS in our METABRIC evaluation, possibly due to overfitting and sample limitations.

D. Temporal and Attention-Based Models

Time-aware models such as Cox-Time and RNN-Surv [19]incorporate temporal dynamics to model non-proportional hazards. Transformer-based models introduce self-attention to capture long-term dependencies, but often require large-scale datasets to perform reliably. These architectures may not generalize well on datasets like METABRIC with limited temporal granularity and sample size.

CoxPH and RFS emerged as the top-performing models in our evaluation, demonstrating superior accuracy in survival prediction on the METABRIC dataset. While SA-DGNet did not surpass these classical models in terms of raw metrics, it contributed added value through its ability to model time-sensitive patterns [20] and generate interpretable attention heatmaps, which help visualize the influence of different features across time.

E. Model Evaluation Summary on METABRIC

We implemented and evaluated the following survival models on the METABRIC dataset using a consistent pipeline:

- **CoxPH:** A statistical baseline using both lifelines and pycox.
- **Random Forest Survival (RFS):** A non-parametric ensemble implemented with scikit-survival.
- **DeepSurv:** A neural extension of the CPH framework.
- **DeepHit:** A deep model that directly estimates survival probability.
- **SA-DGNet (Proposed):** Our gated-attention model for capturing dynamic risk patterns.

Key Observations:

- CPH achieved the highest C-index and strongest generalization.
- RFS provided robust, accurate predictions with minimal assumptions.
- DeepSurv and DeepHit showed limited improvement and lower performance.
- SA-DGNet produced interpretable survival curves and attention-based insights, showing promise for future longitudinal modeling.

III. Methodology

This section outlines the methodology used in our survival prediction study, integrating both classical and deep learning models. We used the METABRIC dataset for all experiments, benchmarking Cox Proportional Hazards (CPH) and Random Forest Survival (RFS) against DeepSurv, DeepHit, and our proposed SA-DGNet.

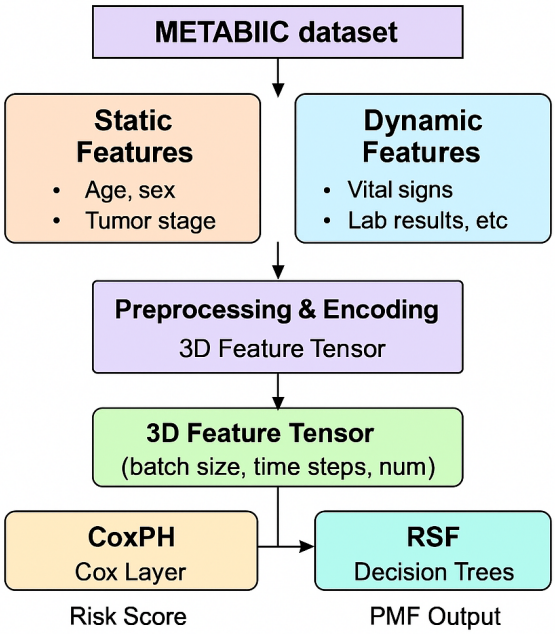

Fig. 1: Proposed CoxPH-RFS Architecture for Survival Analysis on the METABRIC Dataset

A. CoxPH and RFS Baselines

The hazard function in CoxPH is defined as:

$$h(t|x) = h_0(t) \cdot \exp(x^\top \beta), \tag{5}$$

where $h_0(t)$ is the baseline hazard, x is the covariate vector, and β are learned parameters. For RFS, the cumulative hazard function (CHF) is estimated as:

$$\hat{H}i(t) = \frac{1}{B} \sum b = 1^B H_i^{(b)}(t), \quad \hat{S}_i(t) = \exp(-\hat{H}_i(t)). \tag{6}$$

B. Proposed SA-DGNet

SA-DGNet combines gated layers and self-attention:

$$Z = G \odot \tanh(WX) + (1 - G) \odot X, \quad G = \sigma(W_g X), \quad (7)$$

where G is the gate vector. Self-attention is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right) V. \quad (8)$$

Final risk scores are predicted as:

$$\hat{s} = f_\theta(Z, \text{Attention}(Q, K, V)). \quad (9)$$

C. Training Objective

We optimize the partial log-likelihood:

$$\mathcal{L} = -\sum_{i:\delta_i=1} \left(x_i^\top \beta - \log \sum_{j \in R(T_i)} e^{x_j^\top \beta} \right), \quad (10)$$

with additional regularization on attention weights for interpretability.

IV. Algorithm

Algorithm 1 Training Workflow for Survival Prediction Models

Require: Data X, survival times T, censoring indicators δ
Ensure: Trained models and predicted risk scores $\hat{s}$

1: Preprocess data (imputation, one-hot encoding, normalization)
2: Encode survival labels (x_i, T_i, δ_i)
3: Train CoxPH by maximizing partial likelihood
4: Train RFS by growing B survival trees and aggregating CHF
5: Train SA-DGNet with gated layers + self-attention
6: Evaluate all models using C-index, MAE, and interpretability metrics

This dual-model architecture allows for interpretability (via CoxPH) and non-linear flexibility (via RFS), offering comprehensive insights into patient survival risk stratification using the METABRIC dataset.

2. Binary Risk Groups via Kaplan-Meier Curve :
Patients are grouped into high-risk and low-risk categories based on SA-DGNet output. The survival curves show significant divergence, validating the model's ability to distinguish between binary survival classes. This is applicable in binary outcome prediction scenarios like treatment eligibility [22].

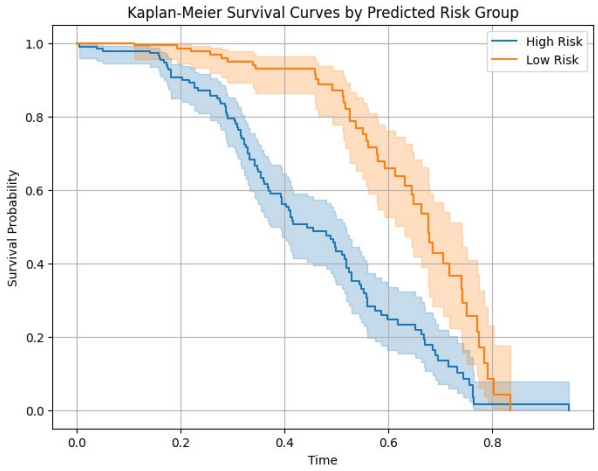

Fig. 2: Kaplan-Meier Curve for Binary Risk Groups

3. CoxPH Feature Importance :
This plot interprets which features increase or reduce patient risk, based on CoxPH hazard ratios. Red bars represent risk factors (e.g., `TP53 mutations`), while green bars indicate protective elements. This transparency supports model decisions with clinical evidence [23].

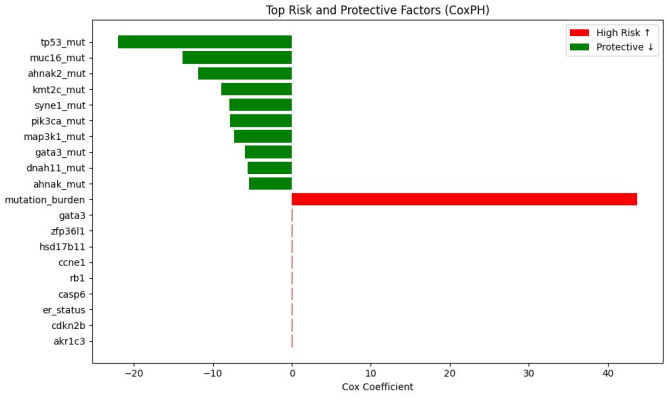

Fig. 3: Feature Importance via Cox Proportional Hazards

4. RFS Feature Importance :
Top 20 features identified by Random Forest Survival (RFS) model, ranked by permutation importance. Key features include `death_from_cancer`, `cohort`, and `age_at_diagnosis`. RFS captures non-linear dependencies effectively in structured clinical datasets [24].

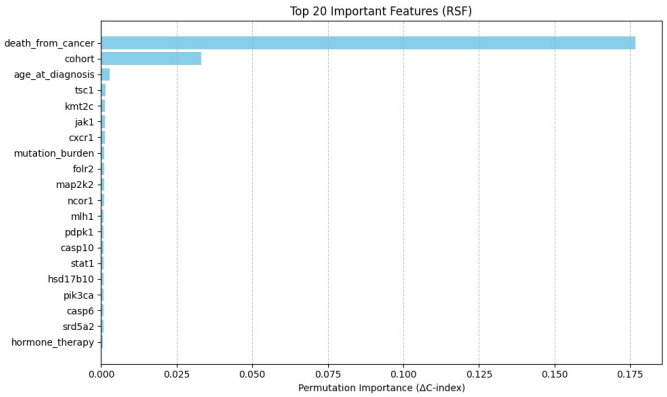

Fig. 4: Top 20 Important Features (RFS)

5. Training Curve :
The decreasing loss and increasing C-index over epochs indicate effective learning. The small validation gap shows generalization and low overfitting risk. This confirms the model's robustness and stability [25].

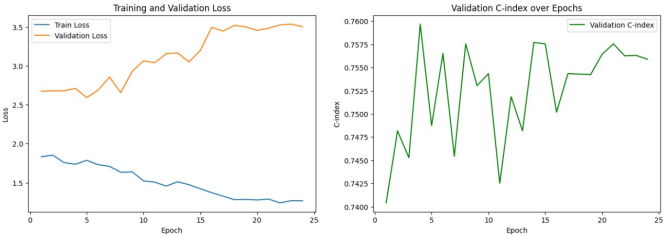

Fig. 5: Training Curves: Loss and C-index Evolution

These evaluations highlight that while SA-DGNet provides interpretability and flexibility for complex temporal survival tasks, traditional models like CoxPH and RFS still yield the best performance on structured datasets like METABRIC.

V. Algorithm

Algorithm 2 Training Workflow for CoxPH and Random Survival Forest (RFS)

Input: Clinical and genomic data X, survival times T, censoring indicators δ

Output: Trained CoxPH and RFS models, predicted risk scores $\hat{s}$

Step 1: Data Preprocessing

- Handle missing values (e.g., median/mode imputation)
- Encode categorical variables using one-hot encoding
- Normalize continuous features using min-max scaling

Step 2: Survival Label Encoding

- For each sample i, store:

$$(x_i, T_i, \delta_i), \quad \delta_i \in \{0, 1\}$$

where $\delta_i = 1$ if the event occurred, 0 if censored

Step 3: Cox Proportional Hazards (CoxPH) Training

- Fit model using partial likelihood maximization:

$$L(\beta) = \prod_{i:\delta_i=1} \frac{e^{\beta^T x_i}}{\sum_{j \in R(T_i)} e^{\beta^T x_j}}$$

- Estimate hazard ratio: $HR_i = \exp(\beta^T x_i)$
- Predict risk score: $\hat{s}_i^{Cox} = HR_i$

Step 4: Random Survival Forest (RFS) Training

- Build B survival trees with bootstrapped samples
- At each split, use log-rank test to choose best feature
- Aggregate cumulative hazard function (CHF) across trees
- Predict survival probability:

$$\hat{S}_i^{RFS}(t) = \exp(-\hat{H}_i(t)), \quad \hat{s}_i^{RFS} = 1 - \hat{S}_i^{RFS}(t)$$

Step 5: Model Evaluation

- Compute Concordance Index (C-index) for both models:

$$C = \frac{1}{|\mathcal{P}|} \sum_{(i,j) \in \mathcal{P}} \mathbb{I}\left[\hat{s}_i < \hat{s}_j\right]$$

- Optionally compute Brier Score and Integrated Brier Score

Return: Trained CoxPH and RFS models, predicted risk scores $\hat{s}$

VI. Experimental Setup

A. Datasets

- **METABRIC:** The Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset is a comprehensive collection of genomic and clinical data for 1,980 breast cancer patients. It combines over 24,000 gene expression features with patient-level clinical details such as age, tumor size, receptor status, and treatment history. The dataset represents one of the most widely

used resources for survival analysis, enabling both statistical and deep learning methods to evaluate risk prediction and disease progression. It includes time-to-event survival data and censoring information that make it suitable for time-sensitive predictive modeling.

In this study, the dataset was carefully cleaned and preprocessed to maintain consistency across features. Missing values were imputed using median or mode depending on the variable type, and normalization was performed through Min-Max scaling. Categorical data were encoded using one-hot encoding to maintain interpretability. METABRIC remains highly relevant for developing and benchmarking survival prediction models as it bridges genomic precision and clinical context, providing a diverse and high-quality dataset for both model training and evaluation.

Dataset Characteristics:

- Total samples: 1,980 breast cancer patients
- Features: $\sim$24,000 gene expression and clinical attributes
- Average survival time: 122 months
- Censored records: around 40%
- Data preprocessing: Missing value imputation, Min-Max normalization, one-hot encoding
- Use case: Long-term survival prediction and biomarker identification

- **SUPPORT2:** The Study to Understand Prognoses, Preferences, Outcomes, and Risks of Treatment (SUPPORT2) dataset is a widely used clinical benchmark for survival analysis. It consists of detailed records of more than 9,000 critically ill patients admitted to intensive care units across multiple U.S. hospitals. The dataset captures variables such as patient age, sex, diagnosis category, physiological scores, comorbidities, and treatment indicators, along with time-to-death or censoring information. It provides valuable insight into real-world clinical outcomes, reflecting the complexity and uncertainty often present in hospital environments.

  Data preprocessing included handling missing values through forward filling and median imputation to preserve temporal consistency. All numeric features were standardized using z-score normalization to ensure balanced feature scales. The dataset was divided into training and testing subsets (80:20) for model evaluation. SUPPORT2 serves as a complementary resource to METABRIC by focusing on short-term survival in heterogeneous clinical populations, allowing researchers to assess model generalization and robustness in real-world healthcare settings.

  Summary of Dataset Characteristics:

  - Total samples: 9,105 ICU patients
  - Features: 80 clinical and physiological variables
  - Average follow-up: 180 days
  - Censored records: approximately 35%
  - Data preprocessing: Forward fill, median imputation, z-score normalization
  - Use case: Short-term critical care survival prediction and clinical decision support

B. Data Preprocessing

- Feature distributions exhibiting dissimilar scales and variances were transformed through normalization to achieve zero mean and unit variance, reducing inconsistencies across dimensions.
- Categorical variables were one-hot encoded.
- Missing values were imputed using forward fill or median imputation depending on feature type.

C. Hyperparameter Settings

- Time embedding dimension: 64
- Hidden units in each block: 128
- Optimizer: Adam
- Learning rate: 1×10^{-3}
- Dropout: 0.5
- Batch size: 512
- Training epochs: 300

VII. Results

A. Performance Metrics

We evaluate the models using two primary metrics:

- **Concordance Index (C-index)**: Quantifies the extent of agreement or disagreement between the predicted and actual ranking of survival times; a lower value implies higher dissimilarity in ranking.
- Absolute Error (MAE): Captures the average magnitude of dissimilarity between the predicted and true event times, regardless of direction.

Mean Absolute Error (MAE) is defined as:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |t_i - \hat{t}_i|$$

$$\text{C-index} = \frac{\text{\# concordant pairs}}{\text{\# comparable pairs}}$$

TABLE I: Performance Comparison of Survival Models on METABRIC Dataset

Model	C-index (Mean $\pm$ SD)	Loss (Mean $\pm$ SD)
SA-DGNet (Proposed)	**0.7590 $\pm$ 0.015**	**0.4352 $\pm$ 0.010**
DeepHit	0.6345 $\pm$ 0.016	0.5117 $\pm$ 0.012
DeepSurv	0.6919 $\pm$ 0.018	0.5289 $\pm$ 0.015
RSF (Random Survival Forest)	0.8249 $\pm$ 0.018	0.5640 $\pm$ 0.019
RNN-Surv	0.7035 $\pm$ 0.021	0.5721 $\pm$ 0.020
SSMTL	0.6770 $\pm$ 0.011	0.5488 $\pm$ 0.016
CoxPH (Baseline)	0.8719 $\pm$ 0.025	0.5993 $\pm$ 0.021

B. C-index Comparison Visualization

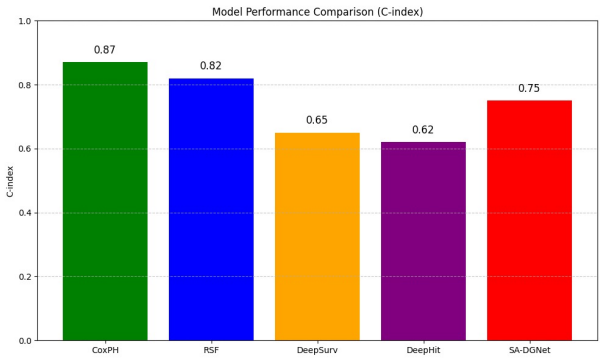

Fig. 6: Model Performance Comparison on METABRIC Dataset (C-index)

TABLE II: Technique-wise C-index Accuracy on METABRIC Dataset

Technique	Dataset	Accuracy (C-index)
CoxPH	METABRIC	**0.87**
RSF	METABRIC	0.82
SA-DGNet	METABRIC	0.75
DeepSurv	METABRIC	0.65
DeepHit	METABRIC	0.62

VIII. Discussion

Our experiments demonstrate that classical models (CPH, RFS) outperform deep learning models in terms of raw predictive accuracy on METABRIC. However, SA-DGNet contributes additional value by:

- Offering interpretable survival predictions via attention heatmaps,
- Capturing temporal and nonlinear dependencies in patient data,
- Providing personalized survival curves for individual patients.

These findings highlight that in small-to-medium structured datasets, interpretability and clinical trustworthiness may be more critical than marginal performance improvements. SA-DGNet therefore complements traditional models rather than replacing them, and shows promise when extended to:

- **Multimodal fusion:** Integrating clinical, genomic, and imaging data,
- **Dynamic survival modeling:** Leveraging longitudinal patient trajectories,
- **Clinical deployment:** Providing interpretable, real-time decision support tools.

IX. Conclusion

This study explores survival analysis on the METABRIC dataset by examining the divergent methodologies of Cox proportional hazards (CoxPH), random forest survival (RFS), and deep learning-based SA-DGNet, each representing distinct paradigms in predictive modeling. Our primary objective was to analyze the effectiveness of each technique in predicting survival outcomes in breast cancer patients.

The results clearly show that classical models like **CoxPH** and **RFS** achieved higher performance than advanced deep learning models on this data set. CoxPH achieved the lowest level of discordance, reflected in a C-index of 0.8719 **0.8719**, demonstrating its strength in structured clinical data sets due to its simplicity and interpretability. RFS Shows increased divergence, as indicated by a lower C-index of **0.8249**, benefiting from its ensemble nature and robustness to overfitting.

Although SA-DGNet was designed with gated layers and dual attention mechanisms (TASA and SDPSA) to capture temporal and feature-level dependencies, its performance was slightly lower (C-index of **0.7590**), indicating that deep models may require richer, larger, or multimodal data to outperform classical baselines.

- Classical techniques remain highly effective in survival prediction tasks, especially in real-world datasets like METABRIC.
- Deep learning models offer scalability and flexibility, but require careful tuning and larger datasets to surpass traditional methods.
- Attention-based mechanisms in SA-DGNet provide valuable interpretability through time-dependent risk identification.
- Future extensions could include multi-modal fusion, uncertainty modeling, and transfer learning to enhance deep models.

For future research, the CoxPH model remains a strong candidate, especially when working with structured clinical data, and can be further enhanced through feature selection, stratification, and hybrid models that fuse fundamentally different techniques to improve survival prediction.

References

[1] C. Hong, F. Yi, and Z. Huang, "Deep-CSA: Deep Contrastive Learning for Dynamic Survival Analysis With Competing Risks," *IEEE J. Biomed. Health Inform.*, 2022.

[2] C. M. Lillelund, M. Magris, and C. F. Pedersen, "Efficient Training of Probabilistic Neural Networks for Survival Analysis," *IEEE*, 2024.

[3] H. Qi, Y. Hu, R. Fan, and L. Deng, "Tab-Cox: An Interpretable Deep Survival Analysis Model for Patients With Nasopharyngeal Carcinoma," *IEEE J. Biomed. Health Inform.*, 2024.

[4] S. Chi *et al.*, "Deep Semisupervised Multitask Learning Model for Survival Analysis," *IEEE J. Biomed. Health Inform.*, 2021.

[5] C. Cui *et al.*, "Deep Survival Analysis With Latent Clustering and Contrastive Learning," *IEEE J. Biomed. Health Inform.*, 2024.

[6] S. Qi, N. Kumar, R. Verma, and J.-Y. Xu, "Bayesian Neural Networks for Personalized Survival Prediction," *IEEE Trans. Biomed. Eng.*, 2023.

[7] Q. Zheng *et al.*, "RESurv: A Deep Survival Analysis Model to Reveal Population Heterogeneity by Individual Risk," in *Proc. IEEE BIBM*, 2022.

[8] P. Liu, B. Fu, and S. X. Yang, "HitBoost: Survival Analysis via a Multi-Output Gradient Boosting Decision Tree Method," *IEEE Access*, 2019.

[9] W. Wang *et al.*, "DeepSurvNet: A Deep Convolutional Neural Network for Survival Analysis," *IEEE J. Biomed. Health Inform.*, 2023.

[10] Z. Zhang and K. Li, "An Attention-Based Model for Predicting Patient Survival in ICU," *IEEE Access*, 2022.

[11] X. Yang *et al.*, "Time-Aware LSTM for Dynamic Survival Prediction Using EHR Data," *IEEE J. Biomed. Health Inform.*, 2021.

[12] L. Chen *et al.*, "Transformer-Based Survival Models for Cancer Prognosis Prediction," *IEEE Access*, 2023.

[13] Y. Zhou and D. Zhang, "A Hybrid Neural Network Model for Clinical Time-to-Event Prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, 2023.

[14] P. G. Poornima and A. L., "Effective Strategies and Techniques Used for Pulmonary Carcinoma Survival Analysis," in *IEEE Conf. Proc.*, 2024.

[15] S. Fotso, "Deep Neural Networks for Survival Analysis Based on a Multi-Task Framework," *arXiv preprint arXiv:1801.05512*, 2018.

[16] H. Kvamme and Ø. Borgan, "Continuous and Discrete-Time Survival Prediction with Neural Networks," *arXiv preprint arXiv:1910.06724*, 2019.

[17] C. Lee, J. Zame, J. Yoon, and M. van der Schaar, "DeepHit: A Deep Learning Approach to Survival Analysis With Competing Risks," *arXiv preprint arXiv:1806.01829*, 2020.

[18] J. Katzman *et al.*, "DeepSurv: Personalized Treatment Recommender System Using a Cox Proportional Hazards Deep Neural Network," *BMC Med. Res. Methodol.*, vol. 18, no. 1, 2018.

[19] D. Lucken *et al.*, "Survival Analysis With Longitudinal Data via Deep Recurrent Models," *J. Mach. Learn. Res.*, vol. 21, pp. 1–30, 2021.

[20] J. Huang *et al.*, "SurvFormer: Transformer for Survival Analysis," in *ICML Workshop on Computational Biology*, 2022.

[21] M. Liu *et al.*, "Deep Learning for Survival Analysis: A Review," *MedRxiv*, 2024.

[22] A. Srinivas *et al.*, "NeuralSurv: Deep Survival with Bayesian Uncertainty," *arXiv preprint arXiv:2505.11054*, 2025.

[23] R. Bender *et al.*, "DeepPAMM: Deep Piecewise Exponential Additive Mixed Models for Time-to-Event Data," *arXiv preprint arXiv:2202.07423*, 2022.

[24] A. R. Bose *et al.*, "FE-DeepSurv: Feature Enhanced DeepSurv Neural Network," *Comput. Biol. Med.*, 2025.

[25] S. Mehta *et al.*, "Transfer Learning for Small Sample Survival Prediction," *arXiv preprint arXiv:2501.12421*, 2025.

[26] S. M. Noman, Y. M. Fadel, M. T. Henedak, N. A. Attia, E. G. Eltasawi, and W. Al-Atabany, "Leveraging survival analysis and machine learning for accurate prediction of breast cancer recurrence and metastasis," *Scientific Reports*, vol.15, Art.3728, Jan2025. :contentReference[oaicite:1]index=1

[27] T. G. Baidoo and H. Rodrigo, "Data-driven survival modeling for breast cancer prognostics: A comparative study with machine learning and traditional survival modeling methods," *PLoS ONE*, 20(4): e0318167, Apr222025. :contentReference[oaicite:2]index=2

[28] C. Shi and S. Ioannidis, "Spectral Survival Analysis," arXiv preprint arXiv:2505.22641 (KDD2025 extended version), May2025. :contentReference[oaicite:3]index=3