Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 7 (2025) 2772

Journal of the Nigerian Society of Physical Sciences

Integrating robust feature selection with deep learning for ultra-high-dimensional survival analysis in renal cell carcinoma

Shaymaa Mohammed Ahmed [a,b,c], Majid Khan Majahar Ali [a,*], Raja Aqib Shamim [a,d]

[a] School of Mathematical Sciences, Universiti Sains Malaysia, 11800, Pulau Penang, Malaysia
[b] Baquba-Technical College, Middle Technical University, Diyala, 32001, Iraq
[c] College of Administration and Economics, University of Diyala, Diyala, 32001, Iraq
[d] Department of Mathematics, University of Kotli, 11100, Azad Jammu and Kashmir, Pakistan

Abstract

The research method applies robust feature selection approaches to ultra-high-dimensional survival data records from Renal Cell Carcinoma patients through deep learning methodologies. The linear methods LASSO and Elastic Net encounter failure when processing data because they face simultaneous multicollinearity issues in addition to overfitting effects and produce marginal survival outcome variability prediction at 54%. We suggest combining ISIS with deep learning architectures featuring PCA-RFA-RSIS models as a remedy to handle these present limitations. Among all evaluated methods PCA-RFA-RSIS is proved most accurate with an MSE measurement of 24.39 and R^2 value of 0.89. PCA improved the model's dimensionality reduction power and robust ISIS maintained model stability despite outliers present in the data. The discovery holds significant value in precision medicine because it creates opportunities to develop individualized therapy for kidney failure patients. Further research needs to enhance hybrid models and expand their utilization between different diseases as well as complex biological systems.

DOI:10.46481/jnsps.2025.2772

Keywords: Ultra-high-dimensional survival analysis, Renal cell carcinoma (RCC), Feature selection with deep learning, Robust SIS, Robust ISIS

Article History :
Received: 18 March 2025
Received in revised form: 13 May 2025
Accepted for publication: 26 May 2025
Published: 15 June 2025

Communicated by: B. J. Falaye

1. Introduction

High-throughput technology advances produced high-dimensional and ultra-high-dimensional (UHD) data, providing strong tools for studying disease-influencing factors, and opening the possibility to study intricate gene-environment-human-action relationships that assist in building patient-specific disease management strategies. The enormous volume of data poses complex challenges to researchers who need advanced analytical methods to find hidden patterns that improve healthcare delivery while advancing precision medical approaches [1].

The statistical method of survival analysis models transitions between states through survival time, which works as a random variable to estimate events, particularly death. Survival data in healthcare mechanisms deals with three main difficulties:

(a) Survival data analysis experiences challenges when the

*Corresponding author Tel. No: +60-14-9543-405.
Email address: majidkhanmajaharali@usm.my (Majid Khan Majahar Ali)

number of predictors (p) exceeds numerous times the number of observations (n) during investigations such as genetic studies ($p > n$) and ultra-high-dimensional conditions ($p > 10^5$).

(b) Meager recording of time-to-event characteristics produces observations that are censored as incomplete data.

(c) Several models depend on the Proportional Hazards Assumption, yet this assumption does not always maintain its validity throughout time-based studies [2].

The analysis becomes more complex when the data volume increases because it introduces additional unnecessary, distorted, and insignificant data. The estimated accuracy declines when the number of model parameters exceeds the number of observations. Model performance suffers from both the effects of multicollinearity and the existence of outliers (such real-world datasets contain 1–10% unexpected data points which differ significantly from the primary dataset) [3].

Standalone survival regression methods fail to handle UHD data sets because they produce inadequate results or inaccurate predictions. Kaplan-Meier analysis and Cox proportional hazards (CPH) models originated when researchers dealt with small datasets, making them unfit for modern high-dimensional or ultra-high-dimensional applications. Modern strategies create new ways to choose risk factors while forecasting survival outcomes [4].

Analyzing complicated variable relationships in risk prediction scenarios benefits from machine learning (ML) techniques, which process substantial complex datasets. Feature selection is a crucial element of ML applications because it provides a means to lower model overfitting, enhancing operational speed and measurement precision. Recent research in machine learning brings new possibilities for analyzing vital risk components through patient survival prediction based on clinical data, despite current implementations of ML for these tasks [5] .

Many regularized approaches exist for selecting variables within high-dimensional data, such as Dantzig Selector, LASSO, Smoothly Clipped Absolute Deviation (SCAD), Adaptive LASSO, and the Minimax Concave Penalty. These methods struggle to work on UHD datasets because the variable count substantially exceeds the number of observations. The techniques exhibit challenges when operating under these conditions because they fail to find a satisfactory number of predictive variables while selecting one variable from related groups of predictors, making LASSO ineffective for UHD applications [6].

Penalized methods such as SCAD and Adaptive LASSO encounter difficulties with ultra-high-dimensional datasets because these data present both statistical precision, computational speed, and algorithm reliability challenges. The Elastic Net (EN) represents an enhanced version of LASSO, which Zou and Hastie developed for managing high correlations between variables. Still, its implementation requires extensive model tuning, which makes analysis of large datasets computationally expensive. The UHD data analysis suffers from implementation difficulties because there is no standardized technique for choosing the λ_1 and λ_2 values [7].

Real-life biological systems made of two components, including kidneys, eyes, and lungs, demonstrate how system components dynamically affect disease development and system stability. Loss of a kidney due to cancer, disease, or injury causes the functioning remaining kidney to face higher failure rates, which increases medical complications. Research demonstrates that these interactions occur in the lung and ear tissue systems [7].

The primary focus of survival analysis models pertains to single-component systems, leading to restricted application with two-component structures. The complex nature of high-dimensional data requires better approaches than traditional stepwise regression because conventional methods, such as stepwise regression and all subsets regression, along with ridge regression, have shown unsatisfactory results. Multiple modern regularized approaches, such as LASSO, Elastic Net, and SCAD, fail to handle the complex nature of UHD survival models in interacting systems [8].

The analysis includes a comparison of specialized variable selection techniques designed particularly for UHD renal cell carcinoma (RCC) survival data because it belongs to a two-component system. Freund's model, which emerged in 1961, is efficient for life testing purposes in biological systems containing two-component hazards. New hybrid models result from Freund's baseline hazard function, deep learning, and iterative sure independence screening approaches. The study analyzes four different analytical approaches: RFA-RISIS (Random Feature Attention with Robust ISIS), Autoencoder-Robust ISIS, Dense-Robust ISIS, as well as PCA-RFA-RISIS (Principal Component Analysis with Random Feature Attention and Robust ISIS) [9].

RCC, along with its 90% dominance among kidney malignancies, serves as the primary case with a specific emphasis on clear cell RCC (KIRC or ccRCC), which represents 70–75% of renal cancers worldwide [10]. This article develops a complete innovative framework for UHD survival analysis in two-component biological systems through the combination of the Freunds' model baseline hazard function and advanced machine learning concepts, including random feature attention, autoencoders, dense networks, and PCA. These novel hybrid approaches achieve efficient identification of vital variables and enhanced assessment of survival patterns in RCC and analogous complex study systems.

2. Related works

Researchers have previously investigated different complex techniques to examine survival patterns and characteristics for renal cell carcinoma and other cancer patients. Medical researchers have applied deep learning analysis of CT images and Cox and Kaplan-Meier statistical methods, alongside which hybrid models demonstrated superior outcomes compared to standard clinical models. The Feature selection algorithms Lasso Freund and Elastic Net Freund operated on genetic data with many features to reveal the genes influencing survival rates.

The single-cell gene expression data analysis received assistance from two deep learning-based models, deep variational transformers and graph neural networks, for therapeutic compound discovery. Evaluation of RAF2Net with deep learning and feature selection resulted in more accurate cancer classification than standard methods because it used deep learning effectively. Interpretable predictive models such as DyS emerged as a solution to combine accurate predictions with the ability to understand the decision-making process. This research applied different predictive models to melanoma and colorectal cancer datasets to show how they achieved effective survival predictions while performing feature selection operations. Deep neural networks utilizing 3D CNN models have enhanced survival prediction accuracy by uniting medical imaging with clinical data. The relevant studies are presented in Table 1.

3. Materials and methods

The examination pursued through Figure 1 depicts our thoughtful workflow design for achieving our goal. The research built a hybrid model for choosing features from ultra-high-dimensional data using the integration process between the Weibull-Freund baseline hazard function and the Cox proportional hazards model to create the Hybrid Cox-Weibull-Freund Model. The UHD RCC dataset underwent processing that split its contents into training at 90% and validation at 10%.

After applying the hybrid model for deep learning-based feature selection approaches. Elastic Net-Robust SIS (EN-Robust SIS) provided process refinement through selection enhancement in two primary hybrid method categories, which included (ENRISIS-Auto ANN), (ENRISIS-Dense ANN), (ENRISIS-RFA), and (ENRISIS-PCA-RFA), as well as (ENRISIS-Auto ANN), (ENRISIS-Dense ANN), (ENRISIS-RFA), and (ENRISIS-PCA-RFA). The evaluation of all eight selected methods focused on MAPE, SSE, MSE, RMSE, and R^2 assessments.

The best selection method emerged from the performance evaluation process. Researchers applied the optimal technique to obtain important features or genes linked to RCC. The final stage incorporated extracted features analysis to improve survival analysis accuracy and interpretation, which used a structured methodology that combined deep learning hybrids to improve RCC patient survival predictions.

3.1. Data description

The 'kidpack' R package provided the obtained gene expression data access. The data frame consists of 4224 DEG records from 74 subject profiles that the system identifies as potential kidney cancers. Many predictors in this dataset qualify it as ultra-high-dimensional data. The investigation included 74 kidney tumor samples exhibiting diverse histological types, differentiation grades, chromosomal abnormalities, and stages, as well as available follow-up data. Patients' survival status and survival rates are present in the provided dataset. A reference was constructed by combining kidney tumor samples to perform hybridization analysis. Researchers can find database-wide information through E-DKFZ-1 < Array Express < Bio Studies < EMBL-EBI and download the data from that location. The study investigates survival hazard as an outcome measure where patient death equals one and survival stands at zero during the research period, using 4224 genes as predictive indicators.

3.2. Data processing

The dataset contained training and validation parts, 90% used for training purposes and 10% for technique validation.

3.3. Cox proportional hazards model

In their studies, survival analysts predominantly use the Cox proportional hazards model to analyze right-censored observations. The model operates with an unspecified independent lifetime distribution and a preset hazard function. The Cox model presents the hazard function as a measure that determines the probability of an event occurring during time t:

$$h(t) = h_0(t) \times \exp(\beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p), \tag{1}$$

or

$$h(t) = h_0(t) \times \exp(x^T \beta). \tag{2}$$

Here, t serves as the time of survival, $h(t)$ represents the function of hazard and $\{x_1, x_2, \ldots, x_p\}$ considered as p covariates values. In contrast, coefficients are denoted as $\{\beta_1, \beta_2, \ldots, \beta_p\}$, which calculate the impact of explanatory factors on the time to survival, and $h_0(t)$ is the unknown baseline hazard function. The unknown parameters are calculated from the partial likelihood by maximizing it.

3.3.1. Weibull Freund- Cox Proportional Hazard Model (WFCPH)

The Hybrid Weibull-Freund-Cox Proportional Hazard (WFCPH) survival model provides essential improvements for analyzing kidney failure duration [11]. The new hybrid model unites Cox proportional hazards properties with all the benefits of the Weibull distribution and the Frechet distribution performance. Analyzing hazard rates over time requires the Weibull component because clinicians need this property when observing risk behavior in kidney failure patients, for example, graft failure and mortality rates evolving at various disease stages [12]. The Freund distribution component adds additional strength to the model's reliable operations [13]:

$$h_{C1}(t_1, t_2 \mid X) = \left(\frac{\beta_{C2}}{\theta_{C2}}\right)\left(\frac{t_2}{\theta_{C2}}\right)^{\beta_{C2}-1}\left(\frac{\beta'_{C1}}{\theta'_{C1}}\right)\left(\frac{t_1-t_2}{\theta'_{C1}}\right)^{\beta'_{C1}-1} \exp(\beta X) \tag{3}$$

The hazard function analysis for the system requires both kidney assessment based on the Weibull-Freund model and the Cox proportional hazards model [14].

$$h(t_1, t_2 \mid X) = h_{C1}(t_1, t_2 \mid X) + h_{C2}(t_1, t_2 \mid X).$$

$$h(t_1, t_2 \mid X) = \left[\begin{array}{l} \left(\frac{\beta_{C2}}{\theta_{C2}}\right)\left(\frac{t_2}{\theta_{C2}}\right)^{\beta_{C2}-1}\left(\frac{\beta'_{C1}}{\theta'_{C1}}\right)\left(\frac{t_1-t_2}{\theta'_{C1}}\right)^{\beta'_{C1}-1} \\ +\left(\frac{\beta_{C1}}{\theta_{C1}}\right)\left(\frac{t_1}{\theta_{C1}}\right)^{\beta_{C1}-1}\left(\frac{\beta'_{C2}}{\theta'_{C2}}\right)\left(\frac{t_1-t_2}{\theta'_{C2}}\right)^{\beta'_{C2}-1} \end{array}\right] \exp(\beta X), \tag{4}$$

Table 1. Summary of studies related to feature selection and prognosis in cancer datasets.

SL No	Authors	Data Source	Methods	Findings
1	Lu et al. [15]	CT images of 822 ccRCC patients	Radiomics + DL; Cox regression; KM analysis	Fusion nomogram outperforms radiomics signature. C-index ↑ by 20% over clinical nomogram.
2	Salma et al. [16]	RCC survival data	Lasso, RLF, SIS, ISIS	RLF-ISIS/SIS outperform others. Identified 49–68 prognostic genes.
3	Chin and Goh [17]	–	rMI-SVM feature selection	Improves classification in high-dimensional data. Optimal features selected efficiently.
4	Ying et al. [18]	–	Deep variational transformer, gradient-based selection	No hyperparam tuning. Deep generative model effective for FS.
5	Kundu et al. [19]	KMC dataset	RAF2Net + TL with ImageNet	Outperforms 9 classifiers. Accuracy, Recall, F-score ¿ 92%.
6	Wang et al. [20]	SKCM dataset	PANDA: Prior-assisted DL + FS layer	PANDA outperforms in FS. Effective on SKCM data.
7	Ness and Udell [21]	–	DyS: Sparse GAM	Discriminative + interpretable for survival data. Suitable for healthcare studies.
8	Wang et al. [22]	GEO scRNA-seq (6 patients)	DL-based compound selection via graph NN	Identified 5 ccRCC-specific drugs using DL on scRNA-seq.
9	Le et al. [23]	TCGA colorectal cancer gene data	Ensemble FS + group Lasso	Better survival prediction. Identifies key high-dim features.
10	Mahootiha et al. [24]	KiTS21 dataset	3D CNN + Spearman + RF + DL survival network	Predicts survival with C-index of 0.84 using CT + clinical data.

Note: FS = Feature Selection, DL = Deep Learning, KM = Kaplan–Meier, TL = Transfer Learning, CNN = Convolutional Neural Network, RF = Random Forest.

Table 2. Performance of LASSO and elastic net as feature selection techniques.

Method	MSE	SSE	RMSE	R^2	Selected Variables	Num.
LASSO	626.5871	8499.337	25.0317	0.4782	X134, X3218, X1728, X3671, X1871, X1866, X2017, X690	8
Elastic Net	625.2725	53600.16	25.9309	0.9063	X134, X3218, X1728, X3671, X1871, X1866, X2017, X166, X690	9

Variables selected by LASSO and Elastic Net. RMSE: Root Mean Square Error, SSE: Sum of Squared Errors, MSE: Mean Squared Error.

where $h_C1(t_1, t_2 \mid X)$ is the hazard function that incorporates two time points for a particular kidney. t_1 and t_2, given a set of covariates $X, \beta_{C1}, \beta_{C2}$. The Weibull distribution relates to the shape parameters that describe kidneys 1 and 2 using the Freund models. Here, θ_{C1}, θ_{C2} are the scale parameters of the Weibull distribution for the Weibull Freund model, corresponding to kidneys 1 and 2, β'_{C1}, β'_{C2} are more shape parameters accompany the model structure to modify or combine variables in the hazard function [25], possibly related to kidney specific factors so $\theta'_{C1}, \theta'_{C2}$ are corresponding scale parameters associated with the additional shape parameters β'_{C1}, β'_{C2} and t_1, t_2 The variable time points indicate particular time intervals during which the hazard functions interact with their coefficients β according to equations (1),(2). [26], Synthesize the covariates' effect with the Weibull Frequentist Cox Proportional Hazard Model structural elements and the dependency treatment by the Frequentist model for two components. [27], where the failure of one affects the failure rate of the other depend on Figure 2, where

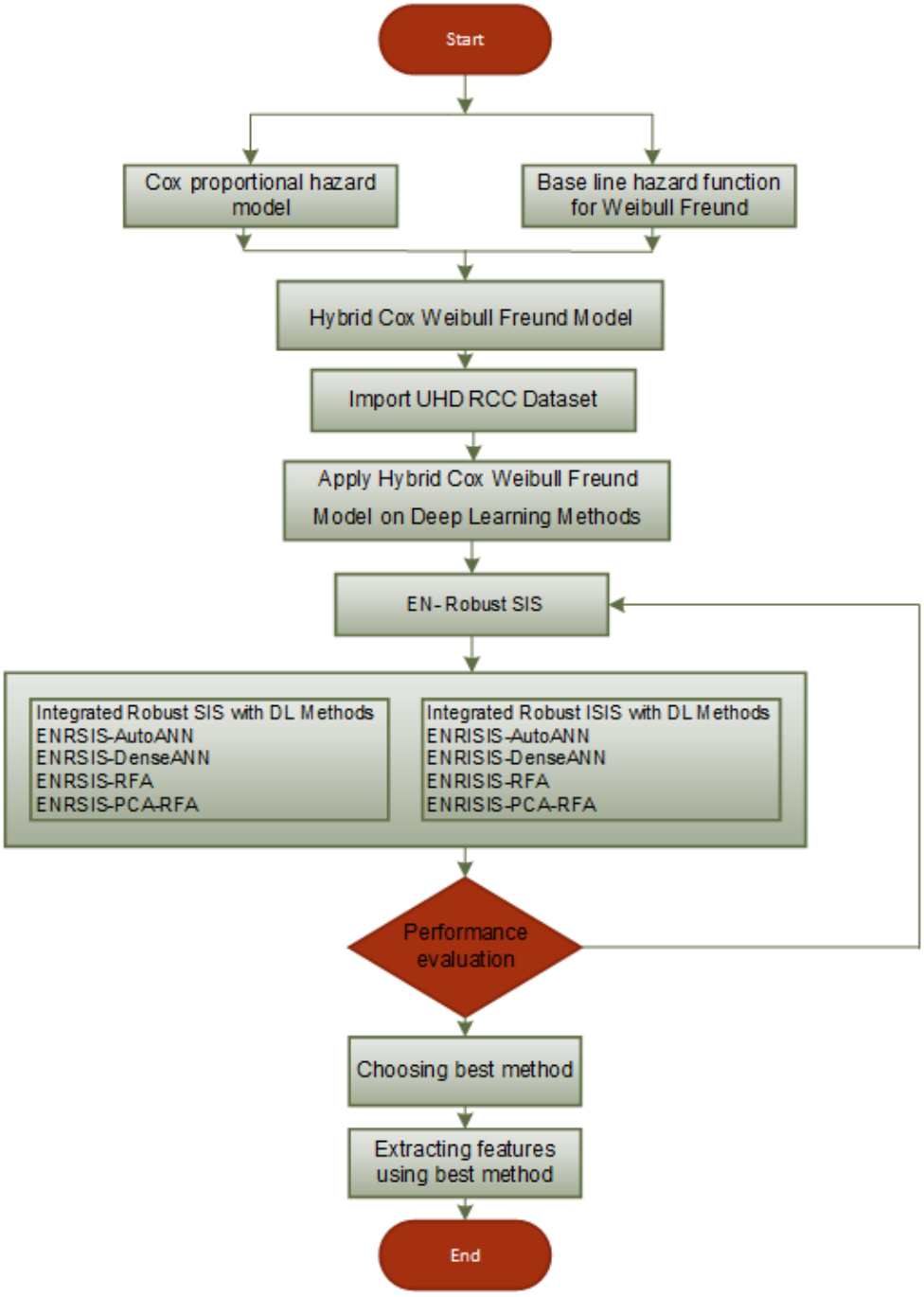

Figure 1. Flowchart of the study's overall methodology.

Dashed Lines represent the baseline hazard rates for components $C1$ and $C2$ before any failure occurs, Solid Lines represent the increased hazard rates after one component fails. For example, the blue solid line shows the increased hazard rate for $C1$ after $C2$ fails, and the green solid line shows the increased hazard rate for $C2$ after $C1$ fails.

Building on this interaction-based hazard modeling, Figure 3 presents the overall architecture of the proposed hybrid survival model. The model integrates both genomic and clinical data to account for such dependent failure behaviors. It begins with gene expression and clinical data preprocessing, followed by feature selection to isolate the most relevant predictors. These selected features are then passed into a hybrid modeling that combines Cox PH, Weibull, and Freund models—each capable of capturing different types of hazard dynamics. This comprehensive structure enables the model to reflect complex dependency patterns, such as those illustrated in Figure 2, and ultimately improves the accuracy of survival predictions.

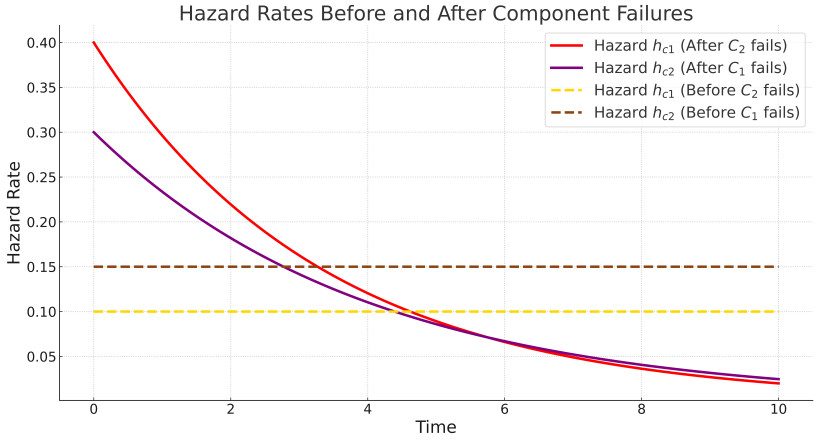

Figure 2. Freund model: hazard functions before and after dependency.

Table 3. Performance of SVM, random forest, and gradient boosting as feature selection techniques.

Method	MSE	SSE	RMSE	R^2	Selected Variables	Num.
SVM	2.3858	19.0864	1.5446	0.9034	x26, x3, x17, x6, x11, x34, x20, x28, x21, x1, x12, x25, x8, x9, x15, x4, x2, x13, x18, x19	20
Random Forest	3.0813	24.6506	1.7554	0.8088	x1, x6, x11, x29, x3, x26, x7, x25, x12, x28, x9, x24, x13, x2, x5, x33, x16, x10, x32, x34	20
Gradient Boosting	2.9036	23.2286	1.7040	0.8733	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 14, 16, 17, 18, 19, 20	20

Variables selected by the respective methods. RMSE: Root Mean Square Error, SSE: Sum of Squared Errors, MSE: Mean Squared Error.

Table 4. Comparison of survival analysis methods with performance metrics for SIS with deep learning.

Methods	MSE	RMSE	SSE	MAE	MAPE	R^2	N.S.G
Autoencoder - RSIS	107.363	10.36162	3972.439	8.23908	6.42465	0.468028	575
Dense - RSIS	92.4763	9.616461	3421.624	7.71147	5.99083	0.550675	620
RFA - RSIS	94.6529	9.728972	3502.157	7.61398	5.29993	0.535903	620
PCA-RFA - RSIS	24.3927	4.938896	902.5296	3.43729	2.81193	0.891191	900

Table 5. Comparison of survival analysis methods with performance metrics for ISIS with deep learning.

Methods	MSE	RMSE	SSE	MAE	MAPE	R^2	N.S.G
RFA-RISIS	105.7942	10.28563	3914.38	7.88155	5.32588	0.485893	240
Autoencoder - RISIS	124.9884	11.17982	4624.57	8.83818	4.67498	0.416952	475
Dense - RISIS	83.52711	9.139317	3090.5	6.62065	3.45732	0.682089	320
PCA_RFA - RISIS	62.00967	7.874622	2294.36	4.32675	1.17373	0.773123	540

3.3.2. Proposed methods

Real-world data analysis for large-scale datasets presents significant difficulties during multi-component system studies, such as kidney, lung, and eye evaluation. There is no established worldwide framework to effectively handle survival data operating across ultra-high dimensions of multi-component systems. This study evaluated multiple specialized variable selection techniques that focus on survival data analysis of high-dimensional multi-component systems to determine efficient methods for these conditions.

Two method groups exist for analysis: conventional approaches and proposed methodologies. The research adopted LASSO with Elastic Net (EN) as one of its selected traditional analytical procedures. The proposed methods include: Lasso Weibull Freund-Sure Independence Screening (Weibull Freund-SIS), Robust Lasso Weibull Freund-Sure Independence Screening (RLF-SIS), Elastic Net Weibull Freund-Sure Independence Screening (ENF-SIS), Elastic Net-Sure Indepen-

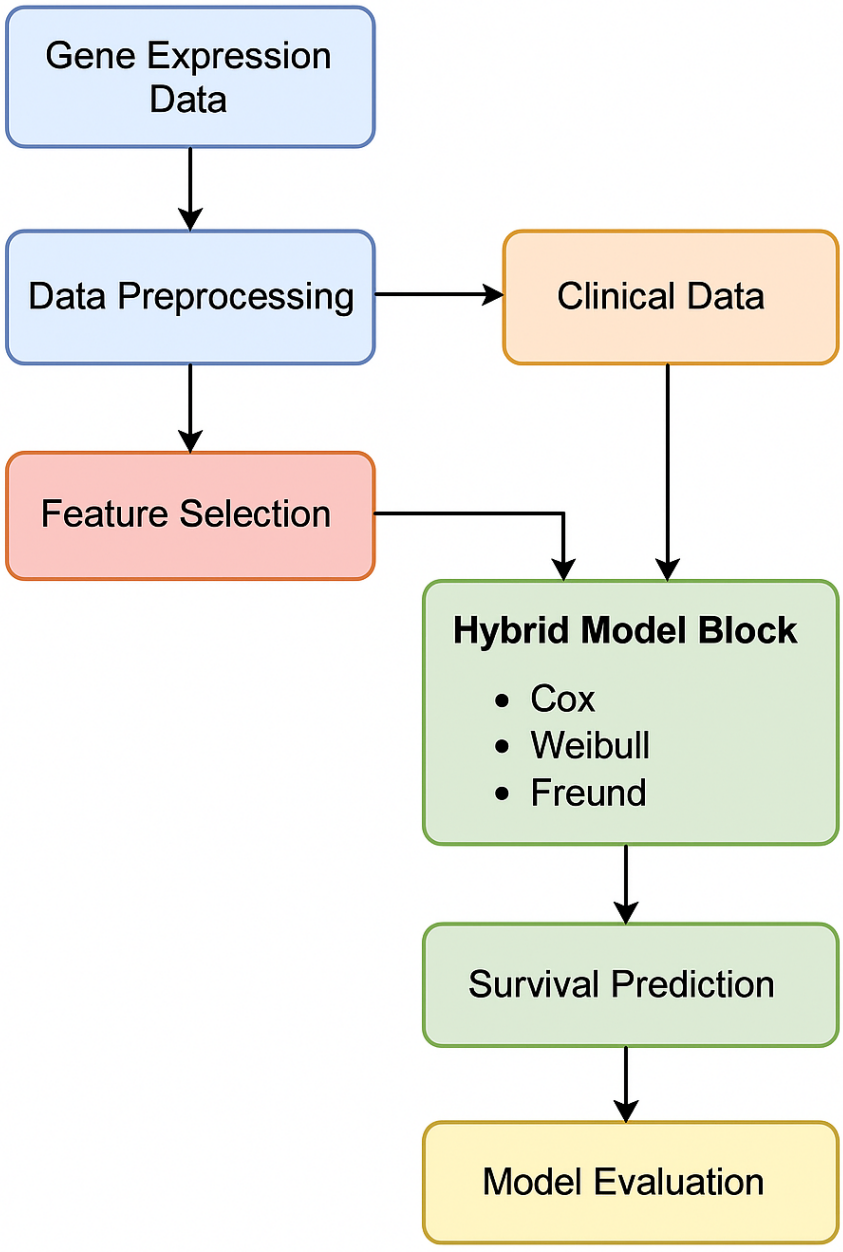

Figure 3. Proposed hybrid model architecture.

dence Screening (EN-SIS), Lasso Weibull Freund-Iterative Sure Independence Screening (Weibull Freund-ISIS), Robust Lasso Weibull Freund-Iterative Sure Independence Screening (RLF-ISIS), Elastic Net-Iterative Sure Independence Screening (EN-ISIS), Elastic Net Weibull Freund-Iterative Sure Independence Screening (ENF-ISIS), RFA-RISIS, Autoencoder-RISIS, Dense-RISIS, and PCA_RFA-RISIS.

Elastic net

The performance capability of LASSO diminishes substantially when the number of variables (p) surpasses the available sample points [28], the sample size restriction limits the number of predictor variables that can be selected in such cases. The LASSO procedure produces limited results when variable groups show strong correlations. The Elastic Net approach by Chamlal *et al.* [7] introduces an improved version of LASSO to handle its existing restrictions. Forecasting accuracy increases through the EN method because it unites the best features of L1 and L2 regularization. The method chooses variables automatically while performing continuous shrinkage procedures. The technique operates similarly to a versatile fishing net, which retains essential covariates (big fish) yet drops unimportant ones. The Elastic Net estimator exists as the following mathematical

definition:

$$J(\beta, \lambda_1, \lambda_2) = \sum_{j=1}^{p} \left[\lambda_1 |\beta_j| + \lambda_2 \beta_j^2\right]. \tag{5}$$

The equation implements Lasso penalties for sparse variables at the beginning, while it applies Ridge penalties for addressing correlated features in the second term to generate more accurate computations. The Elastic Net penalty is a regularization technique suitable for linear models of classification and regression types [1].

$$\hat{\beta}^{\text{elasticnet}} = (1+\lambda_2) \arg\min_{\beta \in \mathbb{R}^p} \left[\begin{array}{l} \frac{1}{N} \sum_{i=1}^{N} (y_i - x_i\beta)^2 \\ +\lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \end{array} \right], \tag{6}$$

where λ_1 and λ_2 are the elastic net's parameters. This method demands finding proper levels of multiple parameters instead of using one individual parameter, and usually adds their values together. The penalty method minimizes the regression loss function for Elastic Net coefficient estimation [29].

$$\sum_{j=1}^{p} \left[\alpha |\beta_j| + (1-\alpha)\beta_j^2\right] \leq \kappa, \tag{7}$$

where κ denotes the extra parameter, which is the total of λ_1 and λ_2.

3.4. Robust Sure Independence Screening (RSIS)

Our approach uses Tukey weights to fortify the robustness of Sure Independence Screening (SIS) when analyzing data affected by outliers and heavy-tailed distributions. The modified screening approaches ensure that the identification and selection procedure becomes less dependent on extreme values, improving variable selection accuracy in high-dimensional applications.

$$\hat{\beta} = \begin{array}{l} \arg\min_{\beta} \left\{ -\sum_{i=1}^{n} w_i \delta_i (x_{(i,M^*)}^T \beta_{(M^*)}) \right. \\ + \sum_{i=1}^{n} w_i \delta_i \log\left(\sum_{j \in R(y_i)} \exp(x_{(j,M^*)}^T \beta_{(M^*)})\right) \\ \left. + p_\lambda(\beta) \right\}, \end{array} \tag{8}$$

where w_i are the Tukey weights computed from the residuals of the initial model, $x_{(i,M^*)}$ is the sub-vector of x_i corresponding to the selected covariates M^*, and $p_\lambda(\beta)$ is a penalty function (e.g., LASSO penalty) [30].

3.4.1. Robust SIS Hybrid with Dense Neural Network (RSIS-DNN)

The study has put forward a combination of Robust Sure Independence Screening (SIS) with Tukey weights and Dense Neural Network (DNN) to handle high-dimensional survival data, including its nonlinear relationships and outliers. This approach benefits from the SIS variable selection strengths combined with the DNN's ability to model complex relationships, making it an appropriate method for survival prediction evaluation [31].

The DNN loss function integrates the Weibull Freund-Cox partial likelihood after its modification with DNN prediction results:

$$L(\theta) = -\sum_{i=1}^{n} \delta_i \left(f_\theta(x_i) - \log\left\{ \sum_{j \in R(y_i)} \exp\left(f_\theta(x_j)\right) \right\} \right), \tag{9}$$

where $f_\theta(x_i)$ is the output of the DNN for the i-th observation, and θ represents the parameters of the DNN [32].

3.4.2. Robust SIS Hybrid with Autoencoder Neural Net work (RSIS-ANN)

The approach combines Robust Sure Independence Screening (SIS) with Tukey weights to address high-dimensional survival data, complex nonlinear relationships, and Autoencoder Neural Networks to determine the Weibull Freund-Cox proportional hazard model. The method unites three key features: robust variable selection, nonlinear feature learning, and survival prediction into a coherent framework.

A neural network generates transformations from d-dimensional covariates to $z \in \mathbb{R}^k$ (where $k < d$) within a latent space, training the autoencoder to minimize reconstruction error using Mean Squared Error [33]:

$$\mathcal{L}_{\text{recon}} = \sum_{i=1}^{n} \|x_i - \text{Decoder}(\text{Encoder}(x_i))\|^2. \tag{10}$$

Using the encoder to transform selected covariates x_i into latent features z_i, we feed z_i into a Weibull Freund-Cox proportional hazard layer to predict the log-hazard function [19]:

$$f_\theta(z_i) = w^T z_i + b. \tag{11}$$

The model includes parameters w and b, using the Weibull Freund-Cox partial likelihood loss with regularization as its loss function:

$$\mathcal{L}_{\text{Cox}} = -\sum_{i=1}^{n} \delta_i \left(f_\theta(z_i) - \log \sum_{j \in R(y_i)} \exp\left(f_\theta(z_j)\right) \right) + \lambda \|w\|^2 . \tag{12}$$

The optimization process jointly handles both the autoencoder and Weibull Freund-Cox Proportional Hazard Model layers:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{Cox}} + \alpha \mathcal{L}_{\text{recon}}, \tag{13}$$

where α balances survival prediction and reconstruction accuracy [19].

3.4.3. Robust Sure Independence Screening - Principal Component Analysis (RSIS- PCA)

Few observations and numerous covariates in survival data create analytical difficulties due to outliers, multicollinearity, and noise [34]. The proposed approach, Robust Sure Independence Screening - Principal Component Analysis (RSIS-PCA), combines robust variable selection techniques with principal component analysis to develop a three-stage method that mitigates the effects of outliers and multicollinearity in survival analysis [23, 35]. The parameter γ is estimated by maximizing the partial likelihood:

$$\mathcal{L}(\gamma) = \prod_{i=1}^{n} \left(\frac{\exp(z_i^T \gamma)}{\sum_{j \in R(y_i)} \exp(z_j^T \gamma)} \right)^{\delta_i}. \tag{14}$$

3.5. Random Feature Attention (RFA)

Random Feature Attention relies upon an unbiased estimation of $\exp(\langle \cdot, \cdot \rangle)$ From Theorem 1 to work properly:

$$\exp\left(\frac{x \cdot y}{\sigma^2}\right) = \exp\left(\frac{\|x\|^2}{2\sigma^2} + \frac{\|y\|^2}{2\sigma^2}\right)\exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \approx \exp\left(\frac{\|x\|^2}{2\sigma^2} + \frac{\|y\|^2}{2\sigma^2}\right)\phi(x)\cdot\phi(y).$$

The last line does not have any nonlinear interaction between $\phi(x)$ and $\phi(y)$, allowing for a linear time/space approximation to attention. We assume the query and keys are unit vectors for clarity[36].

$$\begin{aligned}\text{attn}(q_t, \{k_i\}, \{v_i\}) &= \sum_i \frac{\exp\left(\frac{q_t \cdot k_i}{\sigma^2}\right)}{\sum_j \exp\left(\frac{q_t \cdot k_j}{\sigma^2}\right)} v_i^\top \\ &\approx \sum_i \frac{\phi(q_t)^\top \phi(k_i) v_i^\top}{\sum_j \phi(q_t)\cdot\phi(k_j)} \\ &= \frac{\phi(q_t)^\top \sum_i \phi(k_i)\otimes v_i}{\phi(q_t)\cdot\sum_j \phi(k_j)} = \text{RFA}(q_t, \{k_i\}, \{v_i\}).\end{aligned}$$

$\otimes$ denotes the outer product between vectors, and σ^2 corresponds to the temperature term τ in Eq. 1. RFA can be used as a drop-in replacement for softmax-attention.

(a) The input is fully revealed to cross-attention and encoder self-attention. Here, RFA calculates attention using Eq. 5.

(b) In causal attention, RFA attends only to the prefix.[3] This allows for a recurrent computation. The tuple ($S_t \in \mathbb{R}^{2D\times d}, z_t \in \mathbb{R}^{2D}$) is used as the "hidden state" at time step t to keep track of the history, similar to those in RNNs. Then, RFA $(q_t, \{k_i\}_{i\le t}, \{v_i\}_{i\le t}) = \frac{\phi(q_t)^\top S_t}{\phi(q_t)\cdot z_t}$, where

$$S_t = S_{t-1} + \phi(k_t)\otimes v_t, \quad z_t = z_{t-1} + \phi(k_t). \tag{15}$$

$2D$ denotes the size of $\phi(\cdot)$. Appendix A.1 summarizes the computation procedure of RFA, and Figure 1 compares it against softmax attention. Appendix A.3 derives causal RFA in detail [36].

3.5.1. Robust SIS Hybrid with Random Feature Attention (RSIS-RFA)

We recommend a hybrid method which utilizes Robust Sure Independence Screening (SIS) together with Tukey weights and Random Feature Attention (RFA) for the Weibull Freund-Cox proportional hazard model (SO) to handle high-dimensional survival data features along with outliers and complex nonlinear patterns [25].

The Weibull Freund-Cox partial likelihood loss with regularization is given by:

$$\mathcal{L}_{\text{Cox}} = -\sum_{i=1}^{n}\delta_i\left(f_\theta(\tilde{z}_i) - \log\left(\sum_{j\in R(y_i)}\exp\left(f_\theta(\tilde{z}_j)\right)\right)\right) + \lambda\|w\|^2. \tag{16}$$

The system should optimize its random feature mapping together with its attention system and Weibull Freund-Cox Proportional Hazard Model layer for minimal operation results, $\mathcal{L}_{\text{Cox}}$.

3.6. Robust Iterative Sure Independence Screening (RISIS)

Our proposal involves applying Tukey weight integration to Iterative Sure Independence Screening (ISIS) methods for [30] the Weibull Freund-Cox proportional hazard (WFCPH) model, which enhances resistance to outliers and heavy-tailed data conditions. The modification reduces the sensitivity of the screening and selection processes to extreme observations, increasing the reliability of high-dimensional survival analysis (16). Then, we can write the Penalized Weibull Freund-Cox proportional hazard (WFCPH) model as follows:

Combine the previously selected covariates $\hat{M}_1$ with the newly selected subset L_2 to form an updated set $\hat{M}_2 = \hat{M}_1 \cup L_2$. Estimate the regression coefficients $\hat{\beta}$ using a weighted penalized Weibull Freund-Cox proportional hazard (WFCPH) model:

$$\hat{\beta} = \begin{array}{l}\arg\min_\beta \Big\{-\sum_{i=1}^n w_i\delta_i(x^T_{(i,\hat{M}_2)}\beta_{(\hat{M}_2)}) \\ +\sum_{i=1}^n w_i\delta_i \log\left(\sum_{j\in R(y_i)}\exp(x^T_{(j,\hat{M}_2)}\beta_{(\hat{M}_2)})\right) \\ +p_\lambda(\beta)\Big\},\end{array} \tag{17}$$

where $p_\lambda(\beta)$ is a penalty function.

Also, the refinement process and penalized regression steps should be repeated until a defined stopping criterion, such as multiple iterations or convergence of selected variables, is reached [37].

3.6.1. Robust Iterative Sure Independence Screening Hybrid with Dense Neural Network (RISIS-DNN)

High-dimensional survival data often present challenges such as outliers, noise, and complex nonlinear relationships. We have developed Robust Iterative Sure Independence Screening Hybrid with Dense Neural Network (RISIS-DNN) to resolve these data challenges by integrating robust variable selection and iterative refinement with deep learning technology. The framework utilizes RISIS, which protects against outliers while leveraging DNNs to enhance survival outcome predictions when working with large sets of variables [28, 32].

Using these weights, the robust marginal utility U_m for each covariate is computed as:

$$U_m = \begin{array}{l}\max_{\beta_m}\Big(\sum_{i=1}^n w_i\delta_i x_{im}\beta_m \\ -\sum_{i=1}^n w_i\delta_i \log\left(\sum_{j\in R(y_i)}\exp(x_{jm}\beta_m)\right)\Big),\end{array} \tag{18}$$

where $R(y_i)$ is the risk set at time y_i. Appointment involves the application of a penalized Weibull Freund-Cox Proportional Hazard Model (with LASSO included) to calculate $\hat{\beta}$ using marginal utility evaluation. Screening and refinement methods execute repetitive cycles that end through convergence or attainment of their defined iteration limit, so the chosen variables that display nonzero coefficients undergo updates:

$$\mathcal{L}_{\text{Cox}} = \begin{array}{l}-\sum_{i=1}^n \delta_i\left(f_\theta(x_i) - \log\sum_{j\in R(y_i)}\exp(f_\theta(x_j))\right) \\ +\lambda\|\theta\|^2,\end{array} \tag{19}$$

where θ represents the DNN parameters, and λ serves as the regularization parameter. The DNN obtains its training using gradient-based optimization, while dropout and weight decay are applied to avoid overfitting [16].

3.6.2. Robust Iterative Sure Independence Screening Hybrid with Autoencoder Neural Network (RISIS-ANN)

High-dimensional survival data requires special treatment because it features outliers and signal contamination with noise and unknown nonlinear dependencies. We develop Robust Iterative Sure Independence Screening Hybrid with Autoencoder Neural Network (RISIS-ANN), which merges robust variable selection with iterative refinement and deep learning while resolving these issues. RISIS-ANN unites the resistance to outliers of Robust Iterative Sure Independence Screening with Autoencoder Neural Networks (ANNs) capabilities for feature learning to improve survival predictions among high-dimensional datasets [38], the loss function of Weibull Freund-Cox partial likelihood loss with regularization:

$$\mathcal{L}_{\text{Cox}} = -\sum_{i=1}^{n} \delta_i \left(f_\theta(z_i) - \log \sum_{j \in R(y_i)} \exp(f_\theta(z_j)) \right) + \lambda \|w\|^2, \quad (20)$$

where λ represents the regularization term. The optimization process unites ANN with the Weibull Freund-Cox Proportional Hazard Model through a joint minimization of the combined loss function, with λ serving as the regularization term:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{Cox}} + \alpha \mathcal{L}_{\text{recon}}, \quad (21)$$

where α balances the survival prediction and reconstruction objectives [39].

3.6.3. Robust Iterative Sure Independence Screening Hybrid with Random Feature Attention (RISIS-RFA)

High-dimensional survival data often present challenges such as outliers, noise, and complex nonlinear relationships. The proposed hybrid method, Robust Iterative Sure Independence Screening Hybrid with Random Feature Attention (RISIS-RFA), includes robust variable selection, iterative refinement, and attention mechanisms to solve existing data issues. Survival prediction accuracy in high-dimensional data improves through using RISIS-RFA, which combines the outlier-resistant methods of Robust Iterative Sure Independence Screening (RISIS) with the efficiency from Random Feature Attention (RFA) [21] Then, Weibull Freund-Cox partial likelihood loss with regularization:

$$\mathcal{L}_{\text{Cox}} = -\sum_{i=1}^{n} \delta_i \left(f_\theta(\tilde{z}_i) - \log \left(\sum_{j \in R(y_i)} \exp(f_\theta(\tilde{z}_j)) \right) \right) + \lambda \|w\|^2, \quad (22)$$

where λ is a regularization parameter, the Cox loss is minimized through a joint training process between RFA and the Cox PH model.

3.6.4. Robust Iterative Sure Independence Screening - Principal Component Analysis (RISIS-PCA)

High-dimensional survival data often present challenges such as outliers, noise, and multicollinearity. The proposed framework, Robust Iterative Sure Independence Screening - Principal Component Analysis (RISIS-PCA), utilizes robust variable selection methods alongside iterative refinement while performing dimensionality reduction for these problems. Survival prediction is enhanced in high-dimensional cases through RISIS-PCA because it combines robust variable selection from RISIS with principal component analysis to maintain data variance [40], the Weibull Freund-Cox partial likelihood loss function is used to estimate the parameters:

$$\mathcal{L}_{\text{Cox}} = -\sum_{i=1}^{n} \delta_i \left(f_\theta(z_i) - \log \left(\sum_{j \in R(y_i)} \exp(f_\theta(z_j)) \right) \right) + \lambda \|w\|^2. \quad (23)$$

3.7. Performance Evaluation Criteria

Four metrics were utilized for predictive model accuracy analysis through R^2 coefficient of determination, MSE mean squared error, SSE sum of squares error, and RMSE root mean squared error. Regression tasks form the basis of this analysis because such tasks need metrics that include R^2, MSE, SSE, and RMSE. [41, 42].

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2, \quad (24)$$

where Y_i is the i-th true value, $\hat{Y}_i$ denotes the i-th estimated value, and n represents the overall count of the dataset. An MSE score closer to zero indicates that the model aligns better with the data [43].

The SSE is given by:

$$\text{SSE} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2. \quad (25)$$

In this case, X_i denotes the i-th observation's value, and n is the overall count of the dataset [43]

The RMSE is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2} = \sqrt{\text{MSE}}. \quad (26)$$

The R^2 coefficient is calculated as:

$$R^2 = 1 - \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}. \quad (27)$$

R^2 has two possible values: 0 (the least) and 1 (the largest). To put it in simple terms, a model's R^2 will approach one as the model becomes more successful at predicting outcomes [42].

Also, we used mean absolute percentage error (MAPE), the most frequently utilized summary measure for population forecast error evaluation, and meets all testing criteria[38–43]. MAPE obtains its definition through the following mathematical formula [44] :

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \frac{|\hat{y}_t - y_t|}{y_t} \times 100, \quad (28)$$

where n is the size of the sample, $\hat{y}_t$ is the value predicted by the model for time point t, and y_t is the value observed at time point t.

4. Results and discussion

The interpretation of analysis results is essential in predictive modeling, as it enables the evaluation of model performance and supports data-driven decision-making. The performance evaluation of the proposed models was conducted using a set of statistical measures that assess both prediction accuracy and the ability to explain data variance, such as the mean squared error (MSE) and the coefficient of determination (R^2). All analyses and model implementations were performed using the R programming language (version 4.3.2). The computations were carried out on a workstation equipped with an Intel(R) Core(TM) i7-12700 CPU @ 2.10GHz, 32 GB RAM, running Windows 11. Given the moderate dataset size and model complexity, execution time per model ranged from a few seconds to several minutes. This computational environment was sufficient to ensure efficient training and evaluation of the predictive models.

Table 2 shows a performance analysis between LASSO and Elastic Net at the feature selection task. Elastic Net proved superior to LASSO in data explanation since its coefficient of determination (R^2R^2) achieved 0.9062844 while LASSO only reached 0.4782. Both approaches' prediction accuracy remains similar according to their equivalent RMSE and MSE results. Elastic Net presented an SSE value of 53600.16, which exceeded the LASSO SSE value of 8499.337 due to including the X166 attribute, which increased its model complexity. Elastic Net leads to better variance explanation than LASSO, although LASSO shows stronger capabilities in error reduction for individual predictions.

The selected genes represent a spectrum of biological functions relevant to cellular integrity and signaling. NCAM1 (X134) and CDH8 (X166), both members of the cell-adhesion molecule family, are implicated in maintaining tissue structure and mediating neural growth signals. COL5A1 (X1871) encodes the α-chain of collagen V, contributing to extracellular matrix formation and connective tissue remodeling—functions potentially altered in chronic kidney conditions. ATP1B3 (X3671) is a key component of the Na^+/K^+-ATPase pump, essential for ionic homeostasis and cellular metabolism. The EST corresponding to X1866 exhibits sequence similarity to oxysterol-binding proteins, suggesting a possible involvement in lipid transport and cholesterol regulation. TSC501 (X3218) remains poorly characterized, but its consistent selection indicates potential biological relevance. Finally, unannotated transcripts such as X690, X1728, and X2017 may represent novel regulatory RNAs or yet-to-be-characterized protein-coding genes, offering exciting avenues for future exploration in the context of survival outcomes.

Table 3 shows that feature selection and prediction of SVM delivered superior performance through its highest R^2 value of 0.903, along with minimum RMSE and MSE values, thus demonstrating the best results versus other methods. Gradient Boosting offered performance metrics balanced between the other predictors (R^2=0.873), and Random Forest exhibited slightly lower results (R^2=0.809). Nodes selected by SVM achieve outcomes similar to Elastic Net variants in explaining variance and accuracy, while matching LASSO results, but slightly outperforming them. SVM produces the best results when seeking maximum accuracy combined with variance explanation, but Gradient Boosting is a suitable alternative.

Scientists need specific methods to handle biological data from gene expression studies since these data sets contain extensive dimensions and complex structures. According to the table, the chosen techniques demonstrate efficient feature selection processes during effective management of high-dimensional information. Healthcare practitioners require innovative approaches to improve survival prediction accuracy within their intense professional environment. The selected genes cover a wide array of cellular functions: CDKN2D (X1) encodes $p19^{INK4d}$, a cyclin-dependent kinase inhibitor that regulates cell-cycle checkpoints; FGFR1 (X6) is a receptor tyrosine kinase mediating fibroblast growth factor signaling and cell proliferation; VCAM1 (X11), ICAM1 (X7), and SELE (X28) are key adhesion molecules orchestrating leukocyte–endothelial interactions during inflammation; CTNNB1 (X29) (β-catenin) integrates Wnt signaling with adherens-junction stability; DLG2 (X3) and L1CAM (X32) are neural cell-adhesion proteins crucial for synaptic architecture and neuronal migration; ITGB4 (X26) and LRP1 (X25) serve as transmembrane receptors linking the extracellular matrix to intracellular signaling cascades; JUNB (X9) is an AP-1 transcription factor that modulates gene expression in response to growth and stress signals; CBLB (X12) functions as an E3 ubiquitin ligase attenuating tyrosine-kinase pathways; MCL1 (X24) is an anti-apoptotic Bcl-2 family member safeguarding mitochondrial integrity; KISS1 (X13) acts as a metastasis suppressor in various cancers; EPOR (X5) is the erythropoietin receptor driving erythroid differentiation; GRIN2C (X33) encodes an NMDA-receptor subunit essential for excitatory synaptic transmission; COL18A1 (X16) produces collagen XVIII, a basement-membrane component whose endostatin fragment inhibits angiogenesis; EPLG8 (X10) is a putative ephrin-like ligand modulating cell–cell communication; and ADAM17 (X34) is the TNF-α–converting metalloprotease that sheds cytokine precursors. The EST represented by X2 remains uncharacterized, highlighting an opportunity for future discovery of novel regulators in survival pathways.

4.1. Proposed methods performance

In this section, we will review the results of the proposed methods as follows: Table 4, Autocoder-RSIS and Dense-RSIS, together with RFA-RSIS and PCA-RFA-RSIS, undergo evaluation for predictive performance. The assessment relies on Mean Squared Error (MSE), together with Root Mean Squared Error (RMSE), Sum of Squared Errors (SSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Both the Coefficient of Determination (R^2) and the Number of Selected Genes (N.S.G) evaluate the efficiency of the models.

The Autoencoder-RSIS method produces the most extreme error results, amounting to an MSE of 107.3632, an RMSE of 10.361622, and an SSE of 3972.4386. The predictive accuracy is unsatisfactory when considering the obtained MAE value of 8.23908 and MAPE value of 6.424651. According to the R^2 value of 0.4680276, the survival predictions show only 46.80% variance explained by this methodology. This approach weakens performance in analysing complex high-dimensional gene expression data because of its high MAPE value.

The Dense-RSIS approach provides better results than Autoencoder-RSIS through lower error measurement values. The calculation outcomes demonstrate an MSE value of 92.47632, while the RMSE reaches 9.616461, and the SSE value amounts to 3421.624. The predictive accuracy improves according to the 7.71147 value for MAE and the 5.990825 value for MAPE. The R^2 value rises to 0.5506746, thus offering a 55.07% variance explanation. The higher number of selected genes in the model (N.S.G = 620) enables better performance because it provides a larger set of features for effective data modelling.

RFA-RSIS methodology leads to predictive performance levels identical to Dense-RSIS, although it produces higher error measurement results. The error results from MSE were 94.6529, while the RMSE value stood at 9.728972, and the SSE reached 3502.1572. The model demonstrates better percentage error reduction since it produces an MAPE of 5.299932. The R^2 value reaches 0.5359027 to explain 53.59% of the variance. While producing marginally higher total errors, the enhanced MAPE measurement suggests that RFA-RSIS could provide better outcomes when survival modelling requires minimum relative error assessment. PCA-RFA-RSIS is the most effective method in producing the best error performance indicators. Model performance demonstrates excellence through MSE value 24.39269, RMSE value 4.938896, and SSE value 902.5296. The measurement accuracy of the developed system is supported by the MAE value of 3.43729 and the MAPE value of 2.81193. Among all four methods, the PCA-RFA-RSIS delivers the best explanation of variance since its R^2 value reaches 0.8911911, which translates to 89.12% of explained variance. Implementing Principal Component Analysis (PCA) leads to improved performance because it simplifies dimensions without losing any essential features. The higher number of selected genes (N.S.G = 900) also supports improved model precision.

Although the FA-RSIS algorithm has error metrics that match those of Dense-RSIS, it still demonstrates increased metric values. The method produces MSE results of 94.6529, while RMSE stands at 9.728972 and SSE reaches 3502.1572. The R^2 value for this method amounts to 0.5359027, which indicates a 53.59% variance explanation rate. Among the initial three prediction methods, the FA-RSIS method shows the lowest value of 5.299932 for MAPE, thus suggesting an effective decrease in percentage errors. RFA-RSIS demonstrates enhanced predictive capabilities for survival analysis because it effectively decreases the relative errors found in predictions.

PCA-RFA-RSIS is the most superior technique because it delivers superior prediction accuracy, surpassing every other method. The PCA-RFA-RSIS method establishes optimal results because its error metrics reach 24.39269 MSE, 4.938896 RMSE, and 902.5296 SSE. The prediction accuracy peaks because the model exhibits the lowest MAE value of 3.43729 alongside the lowest MAPE value of 2.81193. This method demonstrates an R-squared value of 0.8911911, which indicates it accounts for 89.12% of survival outcome variance. The performance achievements result from Principal Component Analysis (PCA) integration because it handles dimensionality reduction and preserves key features in the analysis. The predictive model attains better precision and robustness due to the large number of features (N.S.G = 900) selected in this approach.

Table 5 shows that performance evaluation metrics comprising MSE, RMSE, SSE, MAE, MAPE, R^2, and N.S.G. of the methods undergo evaluation to determine their performance in survival function prediction by processing tactical gene expression information.

Autoencoder-RSIS produced the highest errors from all methods when applied to survival data because it yielded an MSE of 107.3632 and an RMSE of 10.361622 alongside an SSE of 3972.4386. The MAE value is 8.23908, and the MAPE value reaches 6.424651, demonstrating limited prediction accuracy. A predictive model with an R-squared value of 0.4680276 clarifies 46.80% of measured changes in the dependent variable. The high MAPE indicates that the method fails to detect crucial patterns embedded within the gene expression database due to its complex and high-dimensional characteristics.

The error metrics from the So Dense-RSIS method surpass the Autoencoder-RSIS metrics, thus proving superior performance. The model performs with a 92.47632 MSE, 9.616461 RMSE, and 3421.624 SSE. The MAE score is 7.71147, while the MAPE shows 5.990825, indicating more accurate predictions. Calculations estimate survival prediction variance at 55.07% through the R^2 value of 0.5506746. Better data feature utilization occurs through the increased number of selected genes (N.S.G = 620), enhancing data modelling accuracy.

5. Conclusion

The research deals with survival analysis problems in renal cell carcinoma (RCC) ultra-high-dimensional (UHD) data through deep learning methods integrated with robust feature selection techniques. Traditional methods, including LASSO and Elastic Net, faced two main difficulties while assessing data with strong relationships and overfitting conditions, which resulted in a 54% explanation of the variance. PCA-RFA-RSIS demonstrated superior predictive accuracy through its performance metrics, such as MSE of 24.39 and R^2 value of 0.89. PCA maintained key information through its ability to simplify data dimensions, and Robust ISIS provided resistance against measurement errors. The research results support new advanced medical treatments since they allow individualized kidney failure treatment strategies. The proposed research requires additional development to extend these models for diverse deployment purposes.

Data availability

No additional data was used beyond those presented in the submitted manuscript.

References

[1] A. Spooner, E. Chen, A. Sowmya, P. Sachdev, N.A. Kochan, J. Trollor & H. Brodaty, "A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction", Scientific reports **10** (2020) 20410. https://doi.org/10.1038/s41598-020-77220-w.

[2] S. Salerno & Y. Li, "High-dimensional survival analysis: Methods and applications", Annual review of statistics and its application **10** (2023) 25. https://doi.org/10.1146/annurev-statistics-032921-022127.

[3] H. Ishwaran, U. B. Kogalur, X. Chen & A. J. Minn, "Random survival forests for high-dimensional data", Statistical Analysis and Data Mining:' The ASA Data Science Journal **4** (2011) 115. https://doi.org/10.1002/sam.10103.

[4] S. Wiegrebe, P. Kopper, R. Sonabend,B. Bischl & A. Bender, "Deep learning for survival analysis: a review", Artificial Intelligence Review **57** (2024) 65. https://doi.org/10.1007/s10462-023-10681-3.

[5] P. Wang, Y. Li & C. K. Reddy, "Machine learning for survival analysis: A survey", ACM Computing Surveys (CSUR) **51** (2019) 1. https://doi.org/10.1145/3214306.

[6] R. Tibshirani, "Regression shrinkage and selection via the lasso", Journal of the Royal Statistical Society Series B: Statistical Methodology **58** (1996) 267. https://doi.org/10.1111/j.1467-9868.2011.00771.x.

[7] H. Chamlal, A. Benzmane & T. Ouaderhman, "Elastic net-based high dimensional data selection for regression", Expert Systems with Applications **244** (2024) 122958. https://doi.org/10.1016/j.eswa.2023.122958.

[8] J. Fan & J. Lv, "Sure independence screening for ultrahigh dimensional feature space", Journal of the Royal Statistical Society Series B: Statistical Methodology **70** (2008) 849. https://doi.org/10.1111/j.1467-9868.2008.00674.x.

[9] R. Reese, X. Dai & G. Fu, "Strong sure screening of ultra-high dimensional categorical data", arXiv preprint (2018) arXiv:1801.03539. https://arxiv.org/abs/1801.07785.

[10] Z. Ba, Y. Xiao, M. He, D. Liu, H. Wang, H. Liang & J. Yuan, "Risk factors for the comorbidity of hypertension and renal cell carcinoma in the cardio-oncologicera and treatment for tumor-induced hypertension", Frontiers in Cardiovascular Medicine **9** (2022) 810262. https://doi.org/10.3389/fcvm.2022.810262.

[11] F. Hamad & N. N. Kachouie, "A hybrid method to estimate the full parametric hazard model", Communications in Statistics-Theory and Methods **48** (2019) 5477. https://doi.org/10.1080/03610926.2018.1513149.

[12] M. Li & B. Ashuri, "Proportional cox hazards model to quantify the likelihood of underestimation in transportation projects", Journal of Construction Engineering and Management **147** (2021) 04021134. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002164.

[13] Z. Y. Li, Q. Shen, J. Tuo, D. D. Tang, L. G. Zhao, and Y. B. Xiang, "Choice and application of time scale selection for Cox proportional hazards regression model in cohort studies", Chin Med J (Engl) **43** (2022) 2002. https://doi.org/10.3760/cma.j.cn112338-20220720-00644.

[14] K. Omae & S. Eguchi, "Quasi-linear cox proportional hazards model with cross-l 1 penalty", BMC Medical Research Methodology **20** (2020) 1. https://doi.org/10.1186/S12874-020-01063-2.

[15] Z. Lu, S. Wu, D. Ni, M. Zhou, T. Wang, X. Zhou, L. Huang & Y. Yan, "Survival analysis of clear cell renal cell carcinoma based on radiomics and deep learning features from ct images", Medicine **103** (2024) 40723. https://doi.org/10.1097/md.0000000000040723.

[16] N. Salma, A. H. M. Al-Rammahi & M. K. M. Ali, "A novel feature selection method for ultra high dimensional survival data", Malaysian Journal of Fundamental and Applied Sciences **20** (2024) 1149. https://doi.org/10.11113/mjfas.v20n5.3665.

[17] F. Y. Chin & Y. K. Goh, "Enhancing classification in high-dimensional data with robust rmi-svm feature selection", Bulletin of Electrical Engineering and Informatics **13** (2024) 3644. https://doi.org/10.11591/eei.v13i5.7938.

[18] W. Ying, D. Wang, H. Chen & Y. Fu, "Feature selection as deep sequential generative learning", ACM Transactions on Knowledge Discovery from Data **18** (2024) 1. https://doi.org/10.1145/3687485.

[19] S. Kundu, N. Roy, R. Talukdar, S. Das, S. Mukhopadhyay & B. Basu Mallik, "Ra f 2net: Automated grading of renal cell carcinoma utilizing attention-enhanced deep learning models through feature fusion", bioRxiv (2024) 2024–07. https://doi.org/10.1101/2024.07.22.604646.

[20] F. Wang, K. Jia & Y. Li, "Integrative deep learning with prior assisted feature selection", Statistics in Medicine **43** (2024) 3792. https://doi.org/10.1002/sim.10148.

[21] M. V. Ness & M. Udell, "Interpretable Prediction and Feature Selection for Survival Analysis", (2024). https://doi.org/10.48550/arxiv.2404.14689.

[22] Y. Wang, X. Chen, N. Tang, M. Guo & D. Ai, "Boosting clear cell renal carcinoma-specific drug discovery using a deep learning algorithm and single-cell analysis", International Journal of Molecular Sciences **25** (2024) 4134. https://doi.org/10.3390/ijms25074134.

[23] P. Le, X. Gong, L. Ung, H. Yang, B. P. Keenan, L. Zhang & T. He, "A robust ensemble feature selection approach to prioritize genes associated with survival outcome in high-dimensional gene expression data", Frontiers in systems biology **4** (2024) 1355595. https://doi.org/10.3389/fsysb.2024.1355595.

[24] M. Mahootiha, H. A. Qadir, J. Bergsland & I. Balasingham, "Multimodal deep learning for personalized renal cell carcinoma prognosis: Integrating ct imaging and clinical data", Computer Methods and Programs in Biomedicine **244** (2024) 107978. https://doi.org/10.48550/arXiv.2307.03575.

[25] H. Ishwaran, U. B. Kogalur, X. Chen & A. J. Minn, "Random survival forests for high-dimensional data", Statistical Analysis and Data Mining: The ASA Data Science Journal **4** (2011) 115. https://doi.org/10.1002/sam.10103.

[26] M. L. Briceno, "A hybrid methodology for the cox proportional hazard model", Economia **33** (2008) 179. https://ideas.repec.org/a/ula/econom/v33y2008i26p179-188.html.

[27] B. Guo & N. Yi, "A scalable and flexible Cox proportional hazards model for highdimensional survival prediction and functional selection", (2022). https://arxiv.org/abs/2205.11600.

[28] J. Jiang & J. Shang, "Feature screening for high-dimensional variable selection in generalized linear models", Entropy **25** (2023) 851. https://doi.org/10.3390/e25060851.

[29] C. M. O'Brien, "Statistical learning with sparsity: the lasso and generalizations", (2016). https://deepblue.lib.umich.edu/bitstream/handle/2027.42/119115/insr12167.pdf?sequence=1.

[30] J. Huang, J. L. Horowitz & F. Wei, "Variable selection in nonparametric additive models", Annals of statistics **38** (2010) 2282. https://doi.org/10.1214/09-AOS781.

[31] Z. Zhang, C.M.M. Padilla, X. Luo, D. Wang & O. H. M. Padilla, "Dense ReLU Neural Networks for Temporal-spatial Model'", (2025). https://arxiv.org/abs/2411.09961.

[32] A.M. Javid, S. Das, M. Skoglund & S. Chatterjee, "A relu dense layer to improve the performance of neural networks", In: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), (2021) 2810. https://doi.org/10.1109/ICASSP39728.2021.9414269.

[33] C. Wang, H. Ma, X. Zhang, X. Xiang, J. Shi, X. Liang, R. Zhao & G. Han, "Deciphering rod pump anomalies: a deep learning autoencoder approach", Processes **12** (2024) 1845. https://doi.org/10.3390/pr12091845.

[34] H. Abdi & L. J. Williams, "Principal component analysis", Wiley interdisciplinary reviews: computational statistics **2** (2010) 433. https://doi.org/10.1002/wics.101.

[35] T. Bankole-Oye, I. El-Thalji & J. Zec, "Combined principal component analysis and proportional hazard model for optimizing condition-based maintenance", IOP Conference Series: Materials Science and Engineering, IOP Publishing **1201** (2021) 012088.https://doi.org/10.1088/1757-899x/1201/1/012088.

[36] H. Peng, N. Pappas, D. Yogatama, R. Schwartz, N. A. Smith & L. Kong, "Random Feature Attention", (2021). http://arxiv.org/abs/2103.02143.

[37] D.R. Cox, "Regression models and life-tables", Journal of the Royal Statistical Society: Series B (Methodological) **34** (1972) 187. https://doi.org/10.1111/j.2517-6161.1972.tb00899.x.

[38] A. H. AL-Rammahi & T. R. Dikheel, "Freund's model with iterated sure

independence screening in cox proportional hazard model", AIP Conference Proceedings, AIP Publishing **2398** (2022) 060009. https://doi.org/10.1063/5.0093464.

[39] Y. Da Wang, M. J. Blunt, R. T. Armstrong & P. Mostaghimi, "Deep learning in pore scale imaging and modeling", Earth-Science Reviews **215** (2021) 103555. https://doi.org/10.1016/j.earscirev.2021.103555.

[40] H. Abdi, L. J. Williams, "Principal component analysis", Wiley interdisciplinary reviews: computational statistics **2** (2010) 433. https://doi.org/10.1002/wics.101.

[41] M. N. Yarahmadi, S. A. MirHassani & F. Hooshmand, "Handling the significance of regression coefficients via optimization", Expert Systems with Applications **238** (2024) 121910. https://doi.org/10.1016/j.eswa.2023.121910.

[42] D. Chicco, M. J. Warrens & G. Jurman, "The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation", Peerj computer science **7** (2021) 623. https://doi.org/10.7717/peerj-cs.623.

[43] L. Zhang, J. Zhang, W. Gao, F. Bai, N. Li & N. Ghadimi, "A deep learning outline aimed at prompt skin cancer detection utilizing gated recurrent unit networks and improved orca predation algorithm", Biomedical Signal Processing and Control **90** (2024) 105858. https://doi.org/10.1016/j.bspc.2023.105858.

[44] C. D. Lewis, "Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting", (1982). https://cir.nii.ac.jp/crid/1130282269656766080.