ISSN: 0067-2904

Unified Machine Learning Techniques for High-Dimensional Survival Data Analysis

Ehab Abbas Fadhil*, Basad Al-Sarray
Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq

Received: 16/6/2023 Accepted: 4/9/2023 Published: 30/11/2024

Abstract

Survival analysis is a statistical method used to analyze time-to-event data. The existence of states where event results become invisible after a given time is one of the primary difficulties with this regulation. This is considered censorship. This kind of data may be found in a variety of applications, such as mechanical system failure rates, patient mortality rates during clinical trials, and the length of unemployment in a community. Assessing the so-called survival and hazard functions is one of the primary goals of survival analysis. Many machine-learning algorithms have also been developed to tackle censored data and other difficult problems with real-world data.

Machine-learning techniques enhance survival analysis by incorporating flexible modeling approaches, handling high-dimensional data, capturing nonlinear relationships, and accounting for censorship. They provide powerful tools to extract meaningful insights and make accurate predictions in time-to-event analysis across various healthcare, finance, and engineering domains. The standard statistical approaches and machine learning techniques advanced for survival analysis are structured in this unified framework, as is the implementation of some machine learning techniques applying to the GBSG2 survival analysis dataset.

Keywords: Deep learning, machine learning, Survival Analysis

تقنيات التعلم الآلي الموحدة لتحليل بيانات البقاء على قيد الحياة عالية الأبعاد

ايهاب عباس فاضل* , بسعاد السراي
قسم علوم الحاسوب ، كلية العلوم ، جامعة بغداد ، بغداد ، العراق

الخلاصة

تحليل البقاء هي طريقة إحصائية تستعمل لتحليل البيانات من وقت إلى حدث معين. وجود حالات التي تصبح فيها نتائج الحدث غير مرئية بعد فترة زمنية معينة هو إحدى الصعوبات الرئيسية في هذه الانظمة. هذا يعتبر رقابة. ويمكن العثور على هذا النوع في مجموعة متنوعة من التطبيقات، مثل معدلات تعطل الجهاز الميكانيكي، ومعدلات وفيات المرضى أثناء التجارب السريرية، وطول فترة البطالة في المجتمع المحلي. وتقييم ما يسمى بوظيفة البقاء والخطر هو أحد الأهداف الرئيسية لتحليل البقاء. كما وُضع العديد من خوارزميات التعلم الآلي لمعالجة البيانات الخاضعة للرقابة وغيرها من المشاكل الصعبة المتعلقة ببيانات العالم الحقيقي.

* Email: Ehab.A.Fadhil@uotechnology.edu.iq

وتعزز تقنيات تعلم الآلات تحليل البقاء على قيد الحياة عن طريق دمج النماذج، ومعالجة البيانات ذات الأبعاد العالية، واستخلاص العلاقات غير الخطية، والحسابات المتعلقة بالرقابة. وهي توفر أدوات قوية لاستخراج أفكار مفيدة وتوقعات دقيقة في التحليل من وقت إلى حدث معين عبر مختلف مجالات الرعاية الصحية والتمويل والهندسة. ويجري في هذا النظام الموحد تنظيم النُهج الإحصائية الموحدة وتقنيات التعلم الآلي المتقدمة لتحليل البقاء على قيد الحياة، وتنفيذ بعض تقنيات التعلم الآلي التي طبقت على مجموعة بيانات تحليل البقاء gbsg2.

1. Introduction

The capacity to gather a large diversity of data and track observations over an extended period of time is now a fact in several dissimilar fields due to the advent of novel data collection and huge data technologies [1]. The essential goal of the popularity of these real-world requests is to more accurately predict the time at which a certain interesting event occurs. One of the core issues with such time-to-event data is censored occurrences, where an important event is missed because of time constraints or becomes lost throughout the observation time, which is very prevalent [2]. Data on survival is often analyzed using statistical and machine-learning techniques. When these issues with censored data develop during the modeling of complicated data, survival analysis, an important part of statistics, offers several strategies to solve them. When simulating a specific event of interest is the primary goal, this is also known as time-to-event data, and it is particularly prevalent in real-world application domains. Several scientists are now developing new computing methods to handle such difficult problems. Machine learning (ML) offers a collection of developed and efficient algorithms that either compete with or complement conventional statistical methods to solve these functional concerns [3]. Recent work has applied machine learning techniques to survival analysis issues. However, despite these issues' significance and applicability to numerous real-world situations, research on this subject is still dispersed. Deep learning (DL) algorithms are a subfield of machine learning algorithms that enable researchers to extract discriminative characteristics with little domain expertise and manual labor. DL has already shown promise in a wide range of areas, including computer vision [4], image compression [5], [6], audio processing, natural language processing [7], robotics, biology and chemistry, medicine [8], video games, search engines, online advertising, detection systems [9], and finance [10].

So, the main goal of this unification is to provide a structured evaluation of several machine learning techniques for survival analysis as well as a discussion of traditional statistical techniques. Show the commonly used estimation and evaluation metrics as well as the complex connected formulations that are often looked into in this area of study. Give a classification of all the survival analysis techniques that have been developed by the machine learning community recently as well as by statisticians in the past, and give a table with all the details about recent studies. The last is the implementation of some machine learning techniques applied to the GBSG2 survival analysis dataset.

The rest of this paper is arranged as follows: Section 2 presents the taxonomy of survival analysis and evaluates the survival probability and regression model used in this study. Section 3 introduces regularized Cox regression. Section 4 is a review of machine learning for survival analysis. Section 5 presents performance evaluation metrics for survival analysis. Section 6 presents the experimental results of machine learning techniques for survival analysis. The last part is the conclusion of this paper.

2. Survival analysis

A subfield of statistics known as survival analysis concentrates on the analysis of time-to-event data, also known as survival times [11]. Data from prospective group studies or data gathered for clinical trials are two examples of data collected prospectively that are typically analyzed using survival analysis methods [12]. It is necessary to specify the time of origin. For instance, the time point of diagnosis for a particular form of cancer might be used as the time origin if the survival time of patients with that kind of cancer is being examined. The last point or event of interest should be adequately characterized so that the times taken into account are clearly understood. In the aforementioned case, cancer research may be to blame for this death. The time taken to go from the time origin to the terminus might then be determined. The presence of scenarios where the outcomes of an event become difficult to observe beyond a certain point in time is one of the main challenges. This is what is referred to as "right censorship." Righteous censorship is the most general type and, therefore, the most straightforward to analyze. When someone is pursued from a time origin (τ_0) up to a late time point (τ_c), right censoring happens since the person hasn't experienced the relevant event. As a result, all that is known about them is that their event hasn't happened as of the censoring time (τ_c). Left censorship is another form of censorship. When a person is left censored, it means that the event occurred before the censoring time; however, it may have happened at any moment in the past. Interval censoring is another option, where a person is only recognized to have experienced the event between two time periods, but the actual event's chosen time remains unrecorded [13]. Events and censorship [14] are shown in Figure-1.

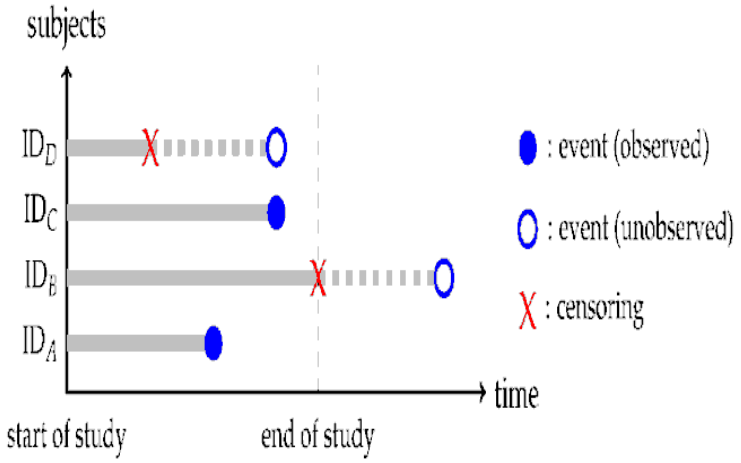

Figure 1: Event and censoring times

2.1. Survival and Hazard Function

The survival function stands for $\zeta(\tau)$, which performs the probability that a person will live at minimal until time t, where $0 \leq \zeta(\tau) \leq 1$. If the survival function is understood through notion or experimental observation, then it may be utilized to comprehend the population's survival experience across time [15].

The survival function has the following properties:

• The definition range of time is $\tau \in [0, \infty)$.

• $\zeta(\tau)$is non-increasing, i.e., $\zeta(\tau 1) \geq \zeta(\tau 2)$ for $\tau 1 \leq \tau 2$

• At time $\tau = 0$, $\zeta(\tau = 0) = 1$, i.e., the probability of surviving past time 0 is 1.

The hazard function $\hbar(\tau)$ is also known as the force of mortality, the instantaneous death rate, or the conditional failure rate. The hazard function indicates the probability of the event happening at the time τ assuming that no event has happened before time τ rather than the chance that the important event will happen [16].

The hazard function $\hbar(\tau)$ has the following properties [17]:

• $\hbar(\tau) \geq 0$ for all τ.
• $\hbar(\tau)$ has no upper bound.
• $\hbar(\tau)$ can assume any shape.

2.2. Taxonomic of Survival Analysis Methods

The main two categories of primary survival analysis techniques are machine learning-based and statistical techniques. Statistical and machine learning techniques aim to predict and evaluate the likelihood of survival at the expected time. Machine learning (ML) techniques are often used to tackle high-dimensional issues, although statistical techniques are typically advanced to deal with low-dimensional data. ML techniques for survival analysis disclose more effective algorithms since they can assess survival problems using both statistics and machine learning methods. As a result, they can benefit from reducing technologies like machine learning and improving them to discover the dependency between covariates and survival times in various pathways [18].

The statistical approaches may be split into three groups: nonparametric, semiparametric, and parametric, which are models depending on the hypothesis made and how parameters are employed in the paradigm. In the last few years, machine learning methods like support vector machines (SVM), neural networks, and survival trees have become increasingly popular. Many have developed machine learning techniques, for example, multitask learning, transfer learning, active learning, and ensemble learning. Figure 2 shows statistical methods for survival analysis.

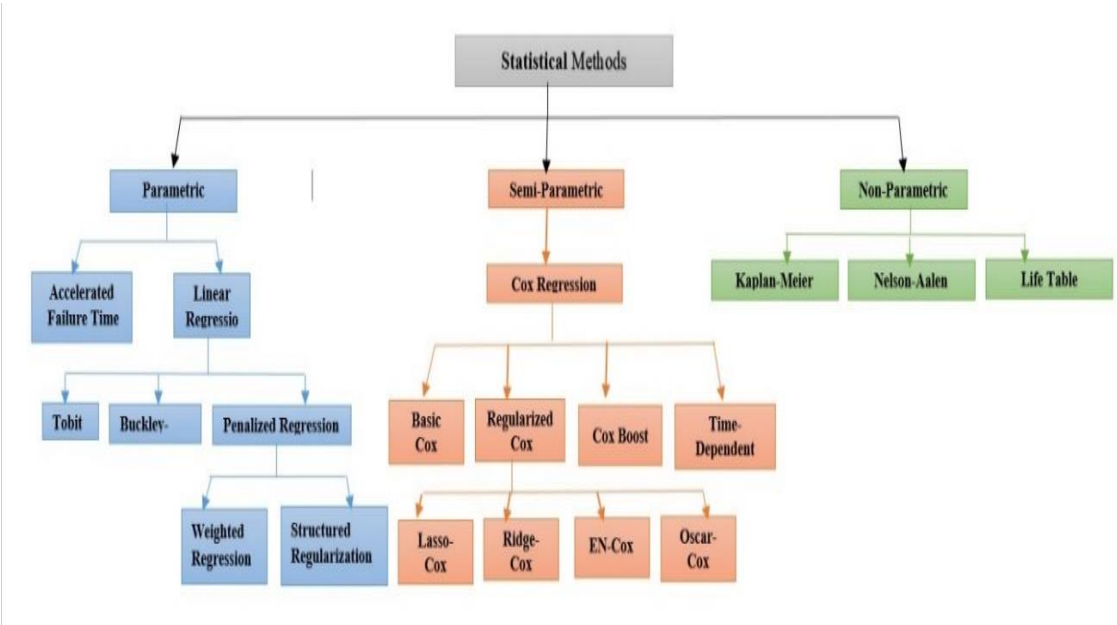

Figure 2: Statistical methods

2.3. Cox Proportional Hazard (CPH) model

A semi-parametric regression model called the CPH model is the most commonly applied. The hazard function is calculated by:

$$\hbar(\tau, \varkappa) = \hbar_0(\tau)\, e^{\beta_i X} \quad (2)$$

The baseline hazard is denoted here as $\hbar 0(\tau)$. The baseline hazard may take on any functional form [19]. The Cox model for survival function can be considered using the formula below.

$$\zeta(\tau) = e^{-H0(\tau)e^{X\beta}} = \zeta 0(\tau) e^{X\beta} \quad (3)$$

- $\mathcal{H}_0(\tau)$ Is the accumulative baseline hazard function.
- $\zeta(\tau) = e^{-\mathcal{H}_0(\tau)}$. Perform the baseline survival function.

Because the baseline hazard function is not specified, the model cannot be fitted with the traditional likelihood function. The Cox model, on the other hand, is free of annoying parameters and applies the partial likelihood function [20], which only depends on the parameter of interest (the baseline hazard). To estimate the coefficient vector, the negative log-partial likelihood is minimized by considering the formula (4):

$$\mathcal{L}(\beta) = -\sum_{j=1}^{N} \delta_j (\mathrm{X}_j \beta \, Log(\sum_{i \in Rj} e^{\mathrm{X}_i \beta})) \quad (4)$$

The numerical Newton-Raphson approach can be used with the Maximum Partial Likelihood Estimator (MPLE) to iteratively discover an estimator $\hat{\beta}$ that minimizes $\mathcal{L}(\beta)$ [21].

2.4. Regularized Cox models

The majority of real-world fields frequently amass high-dimensional data because methods for gathering and analyzing it are always evolving. The number of variables (V) in the given data will sometimes be nearly equal to or higher than the number of instances (I). Therefore, it is a contest to develop the best forecast model that incorporates all of the features available, and in specific instances, the model does yield incorrect data due to overfitting trouble [22], [23]. The related studies for the widely used regularized Cox models are presented in Table 1. Regularized Cox regression is applied in formula (5):

$$\widehat{\beta} = \mathrm{argmin}\beta \; \mathcal{L}(\beta) + \lambda * P(\beta) \quad (5)$$

$P(\beta)$ Is a sparsity-inducing criterion, and λ is the regularization parameter.

Lasso-Cox: The ℓ_1-norm regularizes Lasso is effective at simultaneously doing feature selection and calculating the regression coefficients [24]. Lasso-Cox, considering the formula (6):

$$\widehat{\beta} = \mathrm{argmin}\beta \; \mathcal{L}(\beta) + \lambda * \sum_{v=1}^{v} |\beta_v| \quad (6)$$

Ridge-Cox: A ℓ_2-norm regularization was used in ridge regression to choose the related features and reduce their values toward one another [25] [26]. Ridge-Cox, considering the formula (7):

$$\widehat{\beta} = \mathrm{argmin}\beta \; \mathcal{L}(\beta) + \frac{\lambda}{2} \sum_{v=1}^{v} \beta_v^2 \quad (7)$$

EN-Cox: Elastic net (EN), which mixes ℓ_1 and squared ℓ_2 penalties, can concurrently handle feature selection and feature correlation [27]. Elastic Net-Cox considering the formula (8):

$$\widehat{\beta} = \mathrm{argmin}\beta \; \mathcal{L}(\beta) + \lambda * [\alpha \sum_{v=1}^{v} |\beta_v| + \frac{1}{2}(1+\alpha) \sum_{v=1}^{v} \beta_v^2] \quad (8)$$

OSCAR-Cox: The basic Cox model incorporates the modified graph Octagonal Shrinkage and Clustering Method for Regression (OSCAR) regularization [28] [29] as the OSCAR-Cox algorithm, which may carry out the variable selection for strongly correlated features in regression problems. The key benefit of the OSCAR regularization is that it frequently has equivalent coefficients for features that relate to the result in predictable ways. OSCAR -Cox, considering the formula (9):

$$\widehat{\beta} = \underset{\beta}{\operatorname{argmin}} \mathcal{L}(\beta) + \lambda_1 \| \beta \|_1 + \lambda_2 \| T\beta \|_1 \quad (9)$$

Table 1: Related Studies of Regularized Cox Regression

Authors(Year)	Algorithms
Robert Tibshirani. (1997)[30]	Adaptive Lasso-Cox
Robert Tibshirani, et al. (2005)[31]	Graphical Lasso-Cox
Hao H Zhang and Wenbin Lu, (2007)[32]	Fused Lasso-Cox
Noah Simon, et al. (2011)[33]	Elastic Net-Cox
Jelena Bradic, et al. (2012)[34]	Lasso-Cox
Ying-Wooi Wan, et al. (2013)[35]	Lasso and Elastic Net-Cox
Bhanukiran Vinzamuri and Chandan K Reddy. (2013)[36]	KEN-Cox
Martin Sill,et al. (2014)[37]	Lasso and Elastic Net-Cox
Hokeun Sun, et al. (2015)[38]	Lasso-Cox
Yong Liang, et al. (2016) [39]	L1/2 regularization Cox
Yan Li, et al. (2016)[40]	Transfer-cox
Mansoor Sheikh, *et al. (*2019)[41]	Ridge-Cox
Susana Vinga. (2020)[42]	Elastic Net-Cox
Xinghao Yu, et al. (2020)[43]	Lasso-cox, EN-Cox, mixed-effects Cox model (coxlmm)
RUILIN LI. (2020)[44]	Fast Lasso-cox
J. Kenneth Tay, et al. (2021)[45]	Lasso and Elastic Net-Cox

2.5. Performance Evaluation Metrics

The traditional measures for regression assessment, for instance, the root of mean squared error (MSE) and R^2, are not appropriate for performance evaluation in survival analysis since censoring is present in survival data. For a survival analysis, there are three specific evaluation metrics:

2.5.1 Concordance index (C-index): Its definition is the ratio of concordant pairs to all similar pairs, and it acts as a rank-order statistic for predictions compared to actual outcomes. Assume the comparable instance pair (i, j) with (τ_i) and (τ_j) are the observed times, and $\zeta(\tau i)$ and $\zeta(\tau j)$ are the forecast survival times.

- The pairs(i, j) is concordant if $\tau_i > \tau_j$ and $\zeta(\tau i) > \zeta(\tau j)$.
- The pairs(i, j) is discordant if $\tau_i > \tau_j$ and $\zeta(\tau i) < \zeta(\tau j)$.

Then, the concordance probability:

$$\mathrm{C} = \mathrm{PR}\,(\hat{\tau}_i < \hat{\tau}_j \mid \tau_i < \tau_j) \quad (10)$$

identifies the level of concordance between the ranks of actual values and predicted values [46].

2.5.2 Brier Score: is employed to evaluate forecast models when the output to be predicted is either binary or categorical in form. According to the censoring data, the individual contributions to the empirical Brier score are reweighted:

$$\mathrm{BS}\,(\tau) = \frac{1}{N} + \sum_{i=1}^{N} w_i(\tau)[\,\hat{Y}_i(\tau) - Y_i(\tau)]^2 \quad (11)$$

$w_i(\tau)$ Is used to indicate the weight for the i^{th} instance [47].

2.5.3 Mean Absolute Error: The average of the dissimilarities between the predicted time values and the actual observation time values is the mean absolute error (MAE), which refers to issues with survival analysis.

$$\mathrm{MAE} = \frac{1}{N} + \sum_{i=1}^{N} (\delta_i \mid Y_i - \hat{Y}_i \mid) \quad (12)$$

where Y_i the actual observation times. $\hat{Y}_i$ The predicted times.

In this measure, only the samples for which the event happens are taken into consideration. The mean absolute error may only be used to evaluate survival models that can provide the event time as the forecast target value [48].

3. A review of machine learning for Survival analysis

Significant successes have been made over the past several years in a variety of practical fields due to the benefits of machine-learning techniques. In survival analysis, the main challenges for machine learning algorithms are their ability to model non-linear relationships and the accuracy of their general predictions. Other challenges include how hard it is to work with censored data and how well the model can predict the future. A huge number of instances in comprehensible dimensional feature areas are needed for machine learning to be effective, although this is not always the case for survival analysis issues [49].

The study of algorithms and statistical models that computer systems apply to perform a specific task without being explicitly programmed is known as machine learning (ML). Machine learning is used to train computers to handle data more effectively. At times, after examining the data, it is impossible to explain the information that was extracted from it. In this scenario, machine learning is used. The number of datasets available is increasing the demand for ML. The objective of ML is to learn from the data. Various mathematicians and programmers use a variety of techniques to solve this problem, which makes use of huge amounts of data [50].

Unsupervised machine learning and supervised machine learning are the two basic classifications of machine learning. Unsupervised machine learning, or unsupervised learning, is the method of concluding datasets that contain input data without labeled responses. The objective of supervised machine learning algorithms is to determine the link between input attributes (also known as independent variables) and a target attribute (also known as a dependent variable). Regression and classification are the two basic categories to which supervised techniques belong. While the output variable in classification uses class labels, the output variable in regression uses continuous values [51]. Although there are many other machine learning approaches available, classification is the one that is most frequently employed. In machine learning, classification is a widely respected problem, especially in the areas of knowledge extraction and prediction [52] and [53]. Machine learning methods for survival analysis are shown in Figure 3.

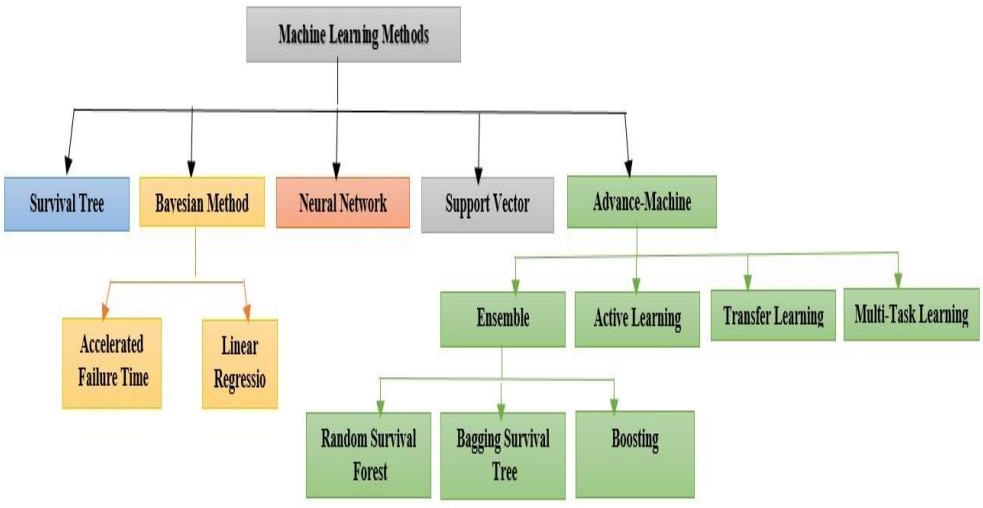

Figure 3: Machine Learning Methods

In many different domains, including survival analysis, the machine learning community has created several highly efficient approaches for high-dimensional settings. The latest studies have extended deep learning with survival analysis methods to select dissimilar data modalities and allow more accurate predictions. For instance, the Deep Convolutional Neural Network (CNN) uses lung cancer data [54]. This study uses diseased images to create a deep convolutional neural network for survival analysis. Asset Health Management Combining deep learning and survival analysis using a dataset of mining haul trucks [55], In order to combine feature extraction and prediction as a single optimization job, this research suggests stacking a three-layer model as a deep learning framework. DeepSurv [56] proposed a method that uses a survival technique based on a Cox proportional hazards deep neural network to model the relationship between a patient's variables and the success of therapy to make personalized treatment recommendations. WSISA [57] presents an efficient methodology for Whole Slide Histopathological Images Survival Analysis (WSISA) to address the aforementioned issues. Cox-nnet [58] is a novel ANN framework developed for predicting patient prognosis from high-throughput transcriptomics data. Comparing Cox-nnet to Cox-proportional hazards regression, random forest survival, and CoxBoost in ten TCGA RNA-Seq data sets using Multi-Task Logistic Regression (MTLR) [59], this study introduces the Neural Multi-Task Logistic Regression (N-MTLR) model as a novel model. Although this model uses MTLR, its deep learning architecture serves as its main core. A scalable discrete-time survival model for neural networks [60] introduces Nnet-survival, a discrete-time survival model that theoretically naturally handles non-proportional hazards and can be trained quickly using mini-batch gradient descent. CXR-risk based on CNN [61] In the present study, a single chest radiograph was used to stratify long-term death risk using the deep learning CXR-risk score. Deep learning-based survival prediction of oral cancer patients [62]: This review used a deep learning-based survival prediction approach for patients with oral squamous cell carcinoma (SCC), confirmed its efficacy, and included 255 patients who were suitable for analysis and had surgical treatment in our department from 2000 to 2017. Dynamic-DeepHit [63] presents a revolutionary deep-learning method that successfully overcomes the current limitations of joint modeling and landmarking by applying their method to a real-world longitudinal dataset from the UK Cystic Fibrosis Registry, which comprises a diverse cohort of 5,883 adult patients with yearly follow-ups between 2009 and 2015. Salmon [64], in this study, uses deep learning-based networks to ascertain the relationship between gene expression data and Cox regression survival in breast cancer. Deep Hazard with a binary variable dataset [65] suggests DeepHazard, a neural network for time-varying hazards, as a novel, flexible strategy for predicting survival. A flexible implementation that supports several optimization techniques without any normative penalty is created. On well-known actual datasets such as METABRIC, GBSG, and ACTG, the same result is shown. In Vale-Silva [66], this research develops an end-to-end discrete-time multimodal DL-based survival prediction technique for estimating patient prognosis. The suggested approach also makes use of a separate network architecture and a distinct data fusion layer. Six input data modalities and 33 different cancer types were combined in MultiSurv. The proposed method, called the Survival Recurrent Network [67], uses all available clinical, pathologic, and therapy data to create a unique deep learning-based survival model (called the Survival Recurrent Network [SRN]) for patients with GC. In this paper, DeepPAMM [68] suggests a new way to handle continuous time-to-event data that works well for many common survival tasks, such as right-censored, left-truncated, competing risks, or multi-state data, as well as recurrent events, the estimation of inherently interpretable feature effects, learning from multiple data sources, time-variy learning, and learning from continuous time-to-event data.

Table 2: Recent studies' explanation

Authors(Year)	Languages	Dataset	Techniques	Results
Xinliang Zhu et.al, (2016)[54]	Python	Pathological Images	CNN (convolutional neural network)	62% with C-Index
Linxia Liao et.al, (2016)[55]	Python	the dataset was collected from a fleet of mining haul trucks	Long Short Term Memory (LSTM) model	training and testing are 87% and 72% with confusion matrix
Jared L. Katzman et.al, (2017)[56]	Python	real clinical data	Cox proportional hazards with deep neural network	95% with C-Index
Safoora Yousefi et.al, (2017)[69]	Python and R package	clinical and molecular data	Comparing Bayesian optimization with Cox elastic net and random survival forests	90% with C-Index
Xinliang Zhu et.al, (2017)[57]	Python	Whole Slide Images (WSIs)	Deep Convolutional Survival CNN and K-mean	70% with C-Index
Travers Ching et.al, (2018)[58]	Python and R package	omics data (10 TCGA)	Cox-nnet (Cox-PH model with the neural network)	85% of C-IPCW scores
Stephane Fotso, (2018)[59]	Python	Worcester Heart Attack Study (WHAS) dataset and (veteran) dataset	Multi-Task Logistic Regression	80% with C-Index
Michael F. Gensheimer et.al, (2019)[60]	Python	MNIST and Real data	nnet-survival using mini-batch stochastic gradient descent (SGD),Cox-nnet and Deepsurv	73% with C-Index
Michael T. Lu et.al, (2019)[61]	Python	Screening from Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) and National Lung Screening Trial (NLST)	CXR-risk based on CNN	PLCO (87%) and NLST (93%) with C-Index
Anika Cheerla et.al, (2019)[70]	Python and shell	clinical, mRNA, microRNA, and WSIs)	CNN (convolutional neural network)	78% with C-Index
Dong Wook Kim et.al, (2019)[62]	Python	Clinical characteristics of the overall dataset	DeepSurv compared with random survival forest (RSF) and CPH model	81% with C-Index
Changhee Lee et.al, (2019)[63]	Python	real-world longitudinal dataset	Dynamic-DeepHit	96% with C-Index
Zhi Huang et.al, (2019)[64]	Python and R package	gene expression (mRNA) and miRNA data	SALMON	72% with C-Index
Havard Kvamme et.al, (2019)[71]	Python	real-world data sets	Cox proportional Hazards (CPH) with (SGD)	79% with C-Index
Lian-Zhen Zhong et.al, (2020)[72]	Python	magnetic resonance (MR) images	multivariable Cox proportional hazards	78% with C-Index
Liwen Zhang et.al, (2020)[73]	R	computed tomography (CT) images	CNN (convolutional neural network)	78% with C-Index
Denise Rava et.al, (2020)[65]	Python	METABRIC, GBSG, and ACTG	Deep Hazard compared with Additive Hazards, Deep Surv,	METABRIC 66% GBSG 68%

		with a binary variable dataset	RSF	ACTG 82% with C Index
Jie Hao et.al, (2020)[74]	Python	histoPAthological Images and Genomic Data (PAGE-Net)	CNN (convolutional neural network)	70% with C-Index
Luís A. Vale-Silva et.al, (2021)[66]	Python and Jupiter Notebook	clinical, imaging, and different high-dimensional omics data modalities	MultiSurv (*CNN*, Adam stochastic gradient descent)	80% with C-Index
Jeeyun Lee et.al, (2022)[67]	Python	clinical and pathologic data	Survival Recurrent Network (SRN)	92% with C-Index
Philipp Kopper et.al, (2022)[68]	Python	tabular and imaging data	DeepPAMM	80% with an integrated Brier score (IBS)

4. Selecting optimization approach and deep learning model

This part discusses the optimization method (**ADMM**) and the deep learning algorithm (**LSTM**) and explains the steps of the two algorithms. They were applied to the survival data set, and the results will be presented in the implementation part.

4.1. Alternating Direction Method of Multipliers (ADMM)

The ADMM optimization approach is commonly used to solve convex optimization issues with decomposable objective functions or constraints [75]. It was developed as a method for effectively handling optimization problems involving both smooth and non-smooth components.

By splitting the larger problem into smaller sub-problems that may be handled separately, the ADMM method is able to solve problems. A sequence of augmented Lagrangian updates alternates between updating the variables related to each sub-problem and taking information from the other variables.

Algorithm 1: pseudocode of ADMM

```
Input: Dataset
Output: Return x
Step 1: Initialize variables:
   Set x, z, u to initial values
   Set step size parameters rho and the maximum number of iterations N
Step 2: Repeat until convergence or maximum iterations are reached:
            For k = 1 to N:
            # Update x
            x = argmin_x L(x, z, u)
            # Update z
            z = argmin_z L(x, z, u)
           # Update u (Lagrange multipliers)
           u = u + rho * (Ax + Bz - c)
           # Check for convergence
           if ||Ax + Bz - c||_2 <= tolerance:
        break
```

The augmented Lagrangian function is represented by L(x, z, u) in the pseudocode above. The updates argmin_x and argmin_z are used to minimize the augmented Lagrangian with respect to x and z, respectively. The constraints of the optimization problem are represented by the matrices and vectors A, B, and c. The dual update step is controlled by the penalty parameter rho, and tolerance is a small value that represents the preferred convergence threshold.

4.2. **The Long Short-Term Memory (LSTM) Model**

LSTM models are part of the deep recurrent neural network family that has been extensively applied in a variety of fields [76]. It was first introduced by Hochreiter and Schmidhuber [77].

LSTMs use a memory cell that can store data over time to understand long-term dependencies in sequential data. Three types of gates—input, forget, and output—control the memory cell. Information entering a cell is controlled by the input gate; information staying in the cell is controlled by the forget gate; and information leaving the cell is controlled by the output gate. Sigmoid activation functions that can accept values between 0 and 1 operate the gates. The LSTM may selectively remember or forget data from earlier time steps because of the sigmoid functions. The gates also take input from the previous and current time steps, allowing the LSTM to learn links between the current and past inputs.

Algorithm 2: LSTM algorithm steps

```
Input: dataset
Output: build model
Step1: Initialize LSTM parameters:
        ( input_size, hidden_size, output_size, learning rate, epochs)
Step2: Initialize LSTM weights and biases
       (Wi, Wf, Wo, Wu, bi, bf, bo, bu)
Step3: Initialize LSTM states
        ht_prev = zeros(hidden_size)
        ct_prev = zeros(hidden_size)
Step4: Training loop
        for epoch in range(epochs):
       # Forward pass
       for t in range(sequence_length):
       xt = input_data[t]
       # Compute activation
       it = sigmoid(W * xt + b + W * ht_prev + b)
       # Compute cell state
       ct = ft * ct_prev + it * ut
      # Compute hidden state
       ht = ot * tanh(ct)
      # Compute output
        output = softmax(Wh * ht + bh)
Step5: # Compute loss
        loss = calculate_loss(output, target_data[t])
```

5. **Experimental Results of machine learning techniques for survival analysis**

Machine learning techniques have been widely applied in survival analysis to improve the prediction and understanding of survival outcomes. Survival analysis deals with time-to-event data, often in the context of studying the time until an event of interest occurs, such as death, disease progression, or the failure of a system.

This implementation will explore several techniques that integrate machine learning with survival analysis. These techniques include the Cox proportional hazards model, the Cox proportional hazards model with penalized estimation, deep learning models, and ADMM. Each technique offers unique advantages and can contribute to improved predictions and understanding of survival outcomes by implementing some techniques and applying machine learning results to the GBSG2 dataset for survival analysis, as shown in Table 3. The GBSG2 (German Breast Cancer Study Group 2) dataset is a well-known benchmark dataset used for survival analysis in machine learning. It contains data on 686 breast cancer patients who

underwent surgery between 1978 and 1982 at the University of Munich. The dataset includes 10 covariates (age, menopausal status, tumor size, etc.) and a response variable indicating whether the patient survived for at least five years following surgery. The dataset was originally used to develop and evaluate prognostic models for breast cancer survival using Cox regression and other survival analysis techniques. The GBSG2 dataset has since been widely used for testing and comparing different machine learning models for survival analysis, including deep learning approaches [78].
The chart of machine learning techniques in survival analysis can be seen in Figure 4.

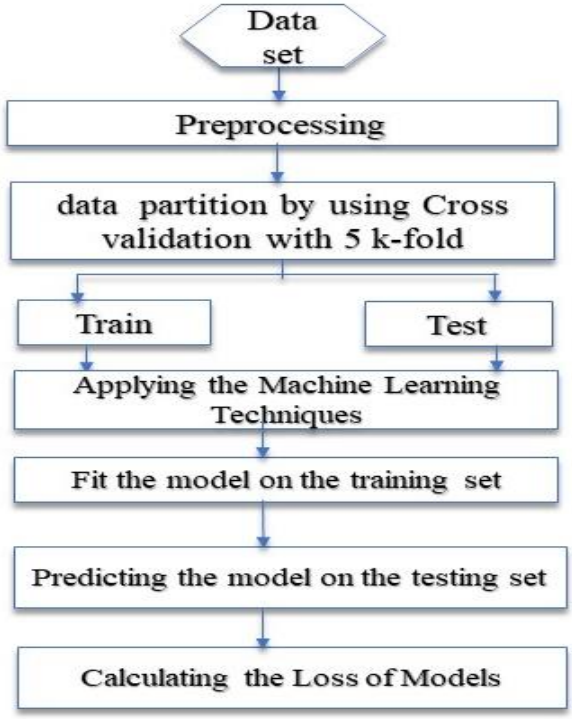

Figure 4: Chart of machine learning techniques in survival analysis

Table 3: Machine Learning Techniques with Survival Analysis

Techniques	Partial Log-Likelihood	Mean Percentage Error (MPE)	Mean Absolute Error (MAE)	C-index
Cox Proportional Hazard model	-8602.907878652857	98.43828556268775	0.5036496350364964	0.691
Lasso Cox model	-8330.533493499	99.97966834953958	0.5452236389040401	0.664
Ridge Cox model	-8463.913708486532	99.94981821978323	0.63719125595535	0.689
Elastic Cox model	-8386.601526220662	99.96368645458143	0.5809251640777783	0.686
ADMM	-8599.444123949035	99.99780980054268	0.7206525066290498	0.691
LSTM	0.1387	99.99568961513981	0.43886545	0.801

Conclusion.

Machine learning techniques have significantly advanced the field of survival analysis, providing powerful tools to analyze time-to-event data and make accurate predictions. These techniques offer several benefits, including handling high-dimensional data, capturing complex relationships, and incorporating censored information.

These techniques have been successfully applied in various healthcare, finance, and engineering domains, enabling researchers and practitioners to gain valuable insights and make informed decisions. Survival analysis models that use machine learning could improve risk prediction, treatment selection, and resource allocation in personalized medicine. They could also help improve survival outcomes in clinical trials and figure out how covariates affect survival times. However, it is important to consider certain challenges and limitations associated with these techniques. These include the need for appropriate feature engineering, addressing bias and overfitting, handling missing data, and interpreting the complex models generated by deep learning approaches. The main objective of this unified text is to offer a structured evaluation of several machine learning techniques for survival analysis, together with a discussion of conventional statistical techniques. Illustrate the widely employed estimation and evaluation metrics as well as the sophisticated connected formulations that are generally explored in linking with this study area, give a classification of all the survival analysis techniques advanced by the machine learning community more recently as well as traditionally by statisticians, and provide a table with all the details about recent studies. The last is the implementation of some machine learning techniques applied to the GBSG2 survival analysis dataset.

References

[1] A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, and S. Belfkih, "Big Data technologies: A survey," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 4, pp. 443-448, 2018.

[2] K.-M. Leung, R. M. Elashoff, and A. A. Afifi, "Censoring issues in survival analysis," *Annual Review of Public Health*, vol. 18, no. 1, pp. 83-104, 1997.

[3] A. Bender, D. Rügamer, F. Scheipl, and B. Bischl, "A general machine learning framework for survival analysis," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2020, Springer, pp. 158-173, 2020.

[4] N. amer Mohamed and M. S. H. Al-Tamimi, "Image fusion techniques: A review," *International Journal of Psychosocial Rehabilitation*, vol. 24, no. 10, pp. 2194-2214, 2020.

[5] A. S. Abd-Alzhra and M. S. H. Al- Tamimi, "Image Compression Using Deep Learning: Methods and Techniques," *Iraqi Journal of Science*, vol. 63, no. 3, pp. 1299–1312, Mar. 2022.

[6] G. Al-Khafaji and M. A. Dagher, "Fixed Predictor Polynomial Coding for Image Compression," *Int. J. Eng. Trends Technol.*, vol. 61, no. 3, p. 182, 2018.

[7] Z. K. . Mohammed and N. A. Abdullah, "Survey For Arabic Part of Speech Tagging based on Machine Learning," *Iraqi Journal of Science*, vol. 63, no. 6, pp. 2676–2685, Jun. 2022.

[8] N. A. K. Hussein and B. Al-Sarray, "Deep Learning and Machine Learning via a Genetic Algorithm to Classify Breast Cancer DNA Data," *Iraqi Journal of Science*, vol. 63, no. 7, pp. 3153–3168, Jul. 2022.

[9] M. A. Hussein, "Performance analysis of different machine learning models for intrusion detection systems," *Journal of Engineering*, vol. 28, no. 5, pp. 61-91, 2022.

[10] S. Pouyanfar et al., "A survey on deep learning: Algorithms, techniques, and applications," *ACM Computing Surveys (CSUR)*, vol. 51, no. 5, pp. 1-36, 2018.

[11] M. Stevenson and I. EpiCentre, "An introduction to survival analysis," EpiCentre, IVABS, Massey University, p.1, 2009.Available: http://www.biecek.pl/statystykaMedyczna/Stevenson_survival_analysis_195.721.pdf

[12] C. Kartsonaki, "Survival analysis," *Diagnostic Histopathology*, vol. 22, no. 7, pp. 263-270, 2016.

[13] S. Prinja, N. Gupta, and R. Verma, "Censoring in clinical trials: review of survival analysis techniques," *Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine*, vol. 35, no. 2, p. 217, 2010.

[14] F. Emmert-Streib and M. Dehmer, "Introduction to survival analysis in practice," *Machine Learning and Knowledge Extraction*, vol. 1, no. 3, pp. 1013-1038, 2019.

[15] L. L. Johnson and J. H. Shih, "An introduction to survival analysis," in *Principles and Practice of Clinical Research: Elsevier*, pp. 273-282, 2007.

[16] O. J. Dunn and V. A. Clark, Basic statistics: a primer for the biomedical sciences. John Wiley & Sons, p.272, 2009.Available:http://www.academia.dk/BiologiskAntropologi/Epidemiologi/PDF/Basic_Statistics__A_Primer_for_the_Biomedical_Sciences_4th_Ed.pdf

[17] T. H. Rashidi and A. Mohammadian, "Parametric hazard functions: overview," *Transportation Research Record*, vol. 2230, no. 1, pp. 48-57, 2011.

[18] Ping Wang, Yan Li, and Chandan K. Reddy, "Machine learning for survival analysis: A survey," ACM Computing Surveys (CSUR), vol. 51, no. 6, p. 110, Nov. 2019.

[19] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187-202, 1972.

[20] D. R. Cox, "Partial likelihood," *Biometrika*, vol 2, p. 269, 1975.

[21] I. H. Hussein and H. A. Hamas, "Fuzzy Survival and Hazard functions estimation for Rayleigh distribution," *Iraqi Journal of Science*, vol. 60, no. 3, pp. 624-632, 2019.

[22] H. van Houwelingen and H. Putter, Dynamic prediction in clinical survival analysis, CRC Press, p.250 2011.

[23] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, The elements of statistical learning: data mining, inference, and prediction, Springer, p. 1, 2009.

[24] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267-288, 1996.

[25] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55-67, 1970.

[26] P. J. Verweij and H. C. Van Houwelingen, "Penalized likelihood in Cox regression," *Statistics in Medicine*, vol. 13, no. 23-24, pp. 2427-2436, 1994.

[27] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301-320, 2005.

[28] S. Yang, L. Yuan, Y.-C. Lai, X. Shen, P. Wonka, and J. Ye, "Feature grouping and selection over an undirected graph," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 922-930, 2012.

[29] J. Ye and J. Liu, "Sparse methods for biomedical data," *ACM Sigkdd Explorations Newsletter*, vol. 14, no. 1, pp. 4-15, 2012.

[30] R. Tibshirani, "The lasso method for variable selection in the Cox model," *Statistics in Medicine*, vol. 16, no. 4, pp. 385-395, 1997.

[31] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 1, pp. 91-108, 2005.

[32] H. H. Zhang and W. Lu, "Adaptive Lasso for Cox's proportional hazards model," *Biometrika*, vol. 94, no. 3, pp. 691-703, 2007.

[33] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for Cox's proportional hazards model via coordinate descent," *Journal of Statistical Software*, vol. 39, no. 5, p. 1, 2011.

[34] J. Bradic, J. Fan, and J. Jiang, "Regularization for Cox's proportional hazards model with NP-dimensionality," *Annals of Statistics*, vol. 39, no. 6, p. 3092, 2011.

[35] Y.-W. Wan, J. Nagorski, G. I. Allen, Z. Li, and Z. Liu, "Identifying cancer biomarkers through a network regularized Cox model," in *2013 IEEE International Workshop on Genomic Signal Processing and Statistics*, IEEE, pp. 36-39, 2013.

[36] B. Vinzamuri and C. K. Reddy, "Cox regression with correlation based regularization for electronic health records," in *2013 IEEE 13th International Conference on Data Mining*, IEEE, pp. 757-766, 2013.

[37] M. Sill, T. Hielscher, N. Becker, and M. Zucknick, "c060: Extended inference with lasso and elastic-net regularized Cox and generalized linear models," *Journal of Statistical Software*, vol. 62, pp. 1-22, 2015.

[38] H. Sun, W. Lin, R. Feng, and H. Li, "Network-regularized high-dimensional Cox regression for analysis of genomic data," *Statistica Sinica*, vol. 24, no. 3, p. 1433, 2014.

[39] Y. Liang, H. Chai, X.-Y. Liu, Z.-B. Xu, H. Zhang, and K.-S. Leung, "Cancer survival analysis using semi-supervised learning method based on Cox and AFT models with L1/2 regularization," *BMC Medical Genomics*, vol. 9, no. 1, pp. 1-11, 2016.

[40] Y. Li, L. Wang, J. Wang, J. Ye, and C. K. Reddy, "Transfer learning for survival analysis via efficient l2, 1-norm regularized cox regression," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*, IEEE, pp. 231-240, 2016.

[41] M. Sheikh and A. C. Coolen, "Analysis of overfitting in the regularized Cox model," *Journal of Physics A: Mathematical and Theoretical*, vol. 52, no. 38, p. 384002, 2019.

[42] S. Vinga, "Structured sparsity regularization for analyzing high-dimensional omics data," *Briefings in Bioinformatics*, vol. 22, no. 1, pp. 77-87, 2021.

[43] X. Yu, T. Wang, S. Huang, and P. Zeng, "How can gene-expression information improve prognostic prediction in TCGA cancers: An empirical comparison study on regularization and mixed Cox models," *Frontiers in Genetics*, vol. 11, p. 920, 2020.

[44] R. Li et al., "Fast Lasso method for large-scale and ultrahigh-dimensional Cox model with applications to UK Biobank," *Biostatistics*, vol. 23, no. 2, pp. 522-540, 2022.

[45] J. K. Tay, B. Narasimhan, and T. Hastie, "Elastic Net Regularization Paths for All Generalized Linear Models," *Journal of Statistical Software*, vol. 106, no. 1, 2023. doi:10.18637/jss.v106.i01

[46] H. Uno, T. Cai, M. J. Pencina, R. B. D'Agostino, and L.-J. Wei, "On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data," *Statistics in Medicine*, vol. 30, no. 10, pp. 1105-1117, 2011.

[47] E. Graf, C. Schmoor, W. Sauerbrei, and M. Schumacher, "Assessment and comparison of prognostic classification schemes for survival data," *Statistics in Medicine*, vol. 18, no. 17-18, pp. 2529-2545, 1999.

[48] S. Hu, E. Fridgeirsson, G. van Wingen, and M. Welling, "Transformer-based deep survival analysis," in *Proceedings of Machine Learning Research*, vol. 146, pp. 132-148, 2021.

[49] B. Zupan, J. Demšar, M. W. Kattan, J. R. Beck, and I. Bratko, "Machine learning for survival analysis: a case study on recurrence of prostate cancer," *Artificial Intelligence in Medicine*, vol. 20, no. 1, pp. 59-75, 2000.

[50] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR)*, vol. 9, no. 1, pp. 381-386, 2020.

[51] A. A. Soofi and A. Awan, "Classification techniques in machine learning: applications and issues," *Journal of Basic & Applied Sciences*, vol. 13, no. 1, pp. 459-465, 2017.

[52] D. Sharma and N. Kumar, "A review on machine learning algorithms, tasks and applications," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 6, no. 10, pp. 2278-1323, 2017.

[53] N. A. S. Aljamali, "Convolutional multi-spike neural network as intelligent system prediction for control systems," *Journal of Engineering*, vol. 26, no. 11, pp. 184-194, 2020.

[54] X. Zhu, J. Yao, and J. Huang, "Deep convolutional neural network for survival analysis with pathological images," in *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, pp. 544-547, 2016.

[55] L. Liao and H.-i. Ahn, "Combining deep learning and survival analysis for asset health management," *International Journal of Prognostics and Health Management*, vol. 7, no. 4, 2016.

[56] J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger, "DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network," *BMC Medical Research Methodology*, vol. 18, no. 1, pp. 1-12, 2018.

[57] X. Zhu, J. Yao, F. Zhu, and J. Huang, "Wsisa: Making survival prediction from whole slide histopathological images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7234-7242, 2017.

[58] T. Ching, X. Zhu, and L. X. Garmire, "Cox-nnet: an artificial neural network method for prognosis prediction of high-throughput omics data," *PLoS Computational Biology*, vol. 14, no. 4, p. e1006076, 2018.

[59] S. Fotso, "Deep neural networks for survival analysis based on a multi-task framework," *arXiv preprint* arXiv:1801.05512, 2018.

Available: https://arxiv.org/pdf/1801.05512.pdf

[60] M. F. Gensheimer and B. Narasimhan, "A scalable discrete-time survival model for neural networks," *PeerJ*, vol. 7, p. e6257, 2019.

[61] M. T. Lu, A. Ivanov, T. Mayrhofer, A. Hosny, H. J. Aerts, and U. Hoffmann, "Deep learning to assess long-term mortality from chest radiographs," *JAMA Network Open*, vol. 2, no. 7, pp. e197416-e197416, 2019.

[62] D. W. Kim, S. Lee, S. Kwon, W. Nam, I.-H. Cha, and H. J. Kim, "Deep learning-based survival prediction of oral cancer patients," *Scientific Reports*, vol. 9, no. 1, p. 6994, 2019.

[63] C. Lee, J. Yoon, and M. Van Der Schaar, "Dynamic-deephit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 1, pp. 122-133, 2019.

[64] Z. Huang et al., "SALMON: survival analysis learning with multi-omics neural networks on breast cancer," *Frontiers in Genetics*, vol. 10, p. 166, 2019.

[65] D. Rava and J. Bradic, "Deephazard: neural network for time-varying risks," *arXiv preprint* arXiv:2007.13218, 2020.

[66] Available: https://arxiv.org/pdf/2007.13218.pdf

[67] L. A. Vale-Silva and K. Rohr, "Long-term cancer survival prediction using multimodal deep learning," *Scientific Reports*, vol. 11, no. 1, p. 13505, 2021.

[68] J. Lee et al., "Deep learning–based survival analysis identified associations between molecular subtype and optimal adjuvant treatment of patients with gastric cancer," *JCO Clinical Cancer Informatics*, vol. 2, pp. 1-14, 2018.

[69] P. Kopper, S. Wiegrebe, B. Bischl, A. Bender, and D. Rügamer, "DeepPAMM: Deep Piecewise Exponential Additive Mixed Models for Complex Hazard Structures in Survival Analysis," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2022: Springer, pp. 249-26, 2022.

[70] S. Yousefi et al., "Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models," *Scientific Reports*, vol. 7, no. 1, p. 11707, 2017.

[71] A. Cheerla and O. Gevaert, "Deep learning with multimodal representation for pancancer prognosis prediction," *Bioinformatics*, vol. 35, no. 14, pp. i446-i454, 2019.

[72] H. Kvamme, Ø. Borgan, and I. Scheel, "Time-to-event prediction with neural networks and Cox regression," *arXiv preprint* arXiv:1907.00825, 2019.
Available: https://www.jmlr.org/papers/volume20/18-424/18-424.pdf?ref=https://githubhelp.com

[73] L.-Z. Zhong et al., "A deep learning MR-based radiomic nomogram may predict survival for nasopharyngeal carcinoma patients with stage T3N1M0," *Radiotherapy and Oncology*, vol. 151, pp. 1-9, 2020.

[74] L. Zhang et al., "A deep learning risk prediction model for overall survival in patients with gastric cancer: A multicenter study," *Radiotherapy and Oncology*, vol. 150, pp. 73-80, 2020.

[75] J. Hao, S. C. Kosaraju, N. Z. Tsaku, D. H. Song, and M. Kang, "PAGE-Net: interpretable and integrative deep learning for survival analysis using histopathological images and genomic data," in *Pacific Symposium on Biocomputing 2020,World Scientific*, pp. 355-366, 2019.

[76] S. Fortunato and C. Castellano, "Community structure in graphs," *arXiv preprint* arXiv:0712.2716, 2007.
Available: https://arxiv.org/pdf/0712.2716.pdf

[77] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting," *Energies*, vol. 13, no. 2, p. 391, 2020.

[78] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.

[79] L. V. Utkin, E. D. Satyukov, and A. V. Konstantinov, "SurvNAM: The machine learning survival model explanation," *Neural Networks*, vol. 147, pp. 81-102, 2022.