



Machine Learning for Time-to-Event Prediction and Survival Clustering: A Review from Statistics to Deep Neural Networks

Jinyuan Luo^{1,3}, Linhai Xie²(✉), Hong Yang¹(✉), Xiaoxia Yin¹,
and Yanchun Zhang^{1,3}(✉)

¹ The Cyberspace Institute of Advanced Technology, Guangzhou University,
Guangzhou, China

2112106041@e.gzhu.edu.cn, {hyang,xiaoxia.yin}@gzhu.edu.cn

² State Key Laboratory of Proteomics, National Center for Protein Sciences
(Beijing), Beijing 102206, China

xielinhai@ncpsb.org.cn

³ School of Computer Science and Technology, Zhejiang Normal University,
Hangzhou, China

Yanchun.Zhang@vu.edu.au

Abstract. Survival analysis is a statistical method used in computational biology to investigate the time until the occurrence of an event of interest, such as death or disease recurrence. It plays a crucial role in analyzing and understanding time-to-event data in various medical and biological studies. Deep learning, as a subset of artificial intelligence, has shown remarkable success in diverse domains, such as image recognition and natural language processing. Its ability to automatically extract complex patterns and features from high-dimensional data makes it highly promising for enhancing survival analysis. While existing reviews have primarily focused on traditional statistical methods and conventional machine learning approaches in survival analysis, a critical aspect that has been largely overlooked is the integration of deep learning techniques. We not only considered traditional statistical methods and conventional machine learning approaches but also further incorporated deep learning methods. With this review article, we hope to provide researchers, doctors, and biologists with a comprehensive framework for understanding survival analysis and to foster the development of survival prediction and personalized medicine.

Keywords: Survival analysis · Time-to-event · Cluster

1 Introduction

Survival analysis is a statistical method used to analyze the probability and influencing factors of survival or experiencing a particular event for individuals

or groups within a given time period [1]. In fields such as medicine, biology, and social sciences, survival analysis methods are widely applied in assessing treatment effectiveness, predicting risks, and studying disease progression, among others [2]. Through survival analysis, doctors and researchers can predict the survival time of patients, develop personalized treatment plans, and evaluate the effectiveness of different treatment approaches based on patient characteristics and clinical data [3]. Additionally, survival analysis can help researchers explore the patterns and pace of disease development, identify genetic variations or biomarkers associated with survival time, and provide scientific evidence for clinical decision-making [4].

However, despite the widespread application of survival analysis methods in various fields, there is relatively limited literature that simultaneously review two different branches of survival analysis, i.e. individualized time-to-event prediction and survival clustering. Except some non-parametric models such as a Kaplan-Meier model, most methods fall into the former branch. They formulate the survival prediction as a regression problem with censored data points and map from the covariates to the estimated risk of an individual, e.g. Cox proportional hazards models. Others, in contrast, are proposed as a unsupervised or semi-supervised clustering task which stratifies patients into subgroups with distinct survival distributions. Although these methods belong to totally different machine learning categories, an integrative overview that comprehensively review both branches of approach can provide deeper insights into survival analysis.

Furthermore, deep learning techniques have shown great promise in enhancing the accuracy and performance of predictive models, especially when dealing with complex and high-dimensional datasets. Its ability to automatically extract relevant features and patterns from raw data has opened up new opportunities for understanding the underlying mechanisms of survival outcomes and identifying critical prognostic factors. Unfortunately, previous reviews and analyses of the literature have provided limited coverage of state-of-the-art deep models in survival analysis.

Therefore, the objective of this review is to systematically review recent advance in survival analysis boosted by deep learning techniques from two different modeling perspectives, including “time-to-event” prediction and survival clustering. We will discuss the applications of statistical methods, traditional machine learning, and deep learning methods in each branch, and explore the advantages, limitations and suitable scenarios for various methods.

The organization of the subsequent sections of this article is as follows: Sect. 2 provides the necessary background knowledge. Section 3 provides an overview of predicting Time to Event. Section 4 introduces risk-based clustering methods. Section 5 concludes the entire article.

2 Preliminaries

In the study of survival analysis problems, it is possible that the event of interest is not observed in certain cases. This can occur due to limited observation time windows or other unobserved events that result in missing information. This concept is referred to as censoring [5]. Censoring can be classified into three groups based on the type of truncation of the event [6]: (I) right censoring, where the observed survival time is less than or equal to the true survival time; (II) left censoring, where the observed survival time is greater than or equal to the true survival time; and (III) interval censoring, where we only know that the event occurred within a given time interval. It is important to note that the true event times are unknown in all three cases. Among these, right censoring is the most common scenario in many practical problems [7]. Therefore, this article will primarily focus on the analysis of survival data with right-censored information.

For survival problems, the time to the event of interest (T) is precisely known only for those instances in which the event occurs during the study period. For the remaining cases, we can only observe a censoring time (C) as we may lose track of them within the observation time or their event occurrence time exceeds the observation time. In the context of survival analysis, they are considered censored instances. In other words, for any given instance i , we can only observe either the survival time (T_i) or the censoring time (C_i), but not both simultaneously. The dataset is referred to as right-censored only when $y_i = \min(T_i; C_i)$ can be observed during the study. In right-censored survival problems, since the samples are entered into the study randomly, and the randomness of the censoring time, the censoring time of the sample is also a random variable. Therefore, in this article, we assume that censoring occurs randomly in survival problems. For brevity, this article refers to randomly occurring right-censored instances as censored instances.

Problem Statement: For a given instance i , it is represented by a triplet $(x_i; y_i; \delta_i)$, where $x_i \in \mathbf{R}^{1 \times P}$ is the feature vector, δ_i is a binary event indicator such that $\delta_i=1$ for uncensored instances and $\delta_i=0$ for censored instances, and y_i represents the observation time, which equals the survival time T_i for uncensored instances and the censoring time C_i for censored instances. Write

$$y_i = \begin{cases} T_i & \text{if } \delta_i = 1 \\ C_i & \text{if } \delta_i = 0 \end{cases} \quad (1)$$

In survival analysis, the objective is to estimate the time from a new instance j , characterized by feature predictors represented by X_j , to an event of interest, T_j . X_j is the variable input to the model. It can be data of various types such as transcriptomics data, medical images, and so on. It is important to note that in survival analysis, the value of T_j is both non-negative and continuous.

3 Time to Event Prediction

Time-to-event prediction, as an application of survival analysis, aims to forecast the probability of a specific event (e.g., death, failure, cure) occurring within a future period. This time frame can be measured from the start of a study or from a particular event's occurrence. Time-to-event prediction holds significant practical significance as it assists in understanding the probability and timing of an event, enabling individuals to make more informed decisions.

3.1 Statistics Based Methods

Statistical methods in survival analysis can be classified into three major categories: (I) parametric methods, (II) non-parametric methods, and (III) semi-parametric methods. Parametric methods are highly effective and accurate in predicting event times when assuming that the data set follows a specific distribution. For instance, if the time in the examined data set follows a well-known theoretical distribution such as the exponential distribution, it is straightforward to use it for estimating event durations. However, in real-life data sets, it is often challenging to obtain data that precisely conforms to known theoretical distributions. In such cases, non-parametric methods can be employed, as they do not rely on assumptions about the underlying distribution of event times. The Kaplan-Meier (Kaplan and Meier, 1958) method is one of the most popular approaches in this category [8]. The Nelson-Aalen (NA) estimator is another non-parametric estimator based on modern counting process techniques [9]. The log-rank test is an application of the Kaplan-Meier method for interval-grouped survival data [10]. The third category comprises a combination of parametric and non-parametric methods. Similar to non-parametric methods, semi-parametric models do not require knowledge of the underlying distribution of event times. The Cox proportional hazards model (Cox, 1972) is the most widely used semi-parametric method in survival analysis [11]. It assumes that the attributes have a multiplicative effect in the hazard function and remain constant over time. In the Cox proportional hazards model, it is assumed that the hazard ratio between two covariates is independent of time and is defined as follows:

$$\lambda(t|x) = \lambda_0(t) \exp(h(x)) \quad (2)$$

$$h(x) = \theta^T x \quad (3)$$

where $\lambda_0(t)$ is the baseline risk. $h(x)$ is the risk function, which represents the impact of covariates on an individual's risk of death. $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ can be estimated using the maximum partial likelihood function. Partial likelihood is the product of the probabilities of individual i being at risk at each event time T_i and experiencing an event. The Cox partial likelihood function is parameterized by θ and defined as follows:

$$L(\theta) = \prod_{i:E=1} \frac{\exp(\hat{h}(x_i))}{\sum_{j \in R(T_i)} \exp(\hat{h}(x_j))} \quad (4)$$

where T_i , E_i , and x_i represent the event time, event indicator, and covariates, respectively, for the i -th observation. The risk set $R(t)$ represents the set of patients who are still at risk of death at time t . The survival function proposed by Cox is highly influential and widely incorporated in subsequent studies within the field of computational biology.

Table 1. Dataset

Dataset	Acronym
Worcester Heart Attack Study	WHAS
Study to Understand Prognoses Preferences Outcomes and Risks of Treatment	SUPPORT
The Molecular Taxonomy of Breast Cancer International Consortium	METABRIC
The Cancer Genome Atlas	TCGA
National Lung Screening Trial	NLST

3.2 Traditional Machine Learning Based Methods

Although statistical techniques aim to characterize the distribution of event times and the statistical properties of each (statistical) model’s parameters, machine learning methods seek to predict event occurrences at given time points. The decision tree algorithm (Bou-Hamad et al., 2011) is based on recursive partitioning with specific splitting criteria applicable to survival analysis [12]. Due to the key feature of this algorithm being the splitting criterion, there has been some research focused on finding effective splitting criteria for survival analysis [13]. Random survival forests (RSF) employ the log-rank test as the splitting criterion to construct random forests. It calculates the cumulative hazards of leaf nodes and averages them over all elements [14]. The LASSO-COX model, utilizing the least absolute shrinkage and selection operator (LASSO), applies feature selection to choose relevant subsets for cancer prediction [15]. SVRc improves the loss function to handle truncated data [16]. It leverages the advantages of standard Support Vector Regression (SVR) and makes it applicable to censored cases through an updated asymmetric loss function that considers both uncensored and censored instances in the model.

Bayesian analysis is one of the fundamental principles in statistics, linking posterior probabilities with prior probabilities. Some studies employ this model to predict the probability of the events of interest [17], benefiting from the desirable properties of Bayesian modeling, such as interpretability [18]. Features in Bayesian networks can be interrelated at different levels and can be represented graphically to depict the theoretical distribution of a set of variables. Bayesian networks provide an intuitive representation of all relationships among variables, making them interpretable for end-users. Knowledge information can be acquired by estimating the network structure and parameters from a given dataset. (Fard

et al., 2016) introduced a novel framework that combines the representational power of Bayesian networks with the accelerated failure time (AFT) model by extrapolating the prior probabilities to future time points [19]. The computational complexity of these Bayesian methods primarily depends on the type of Bayesian techniques used in the model.

Support Vector Machine (SVM) is also a crucial class of machine learning algorithm [20]. It can be used for both classification and regression tasks and has been successfully applied to survival analysis problems [21]. (Van et al., 2007) studied a learning machine designed for predictive modeling of survival data with independent correct censoring by introducing a health index as a proxy between instance covariates and outcomes [22]. (Van et al., 2011) presented an SVR-based method that combines ranking and regression approaches within the context of survival analysis [23]. On average, these methods have a time complexity of $O(N^3)$, which is comparable to the time complexity of standard Support Vector Machines.

3.3 Neural Network Based Methods

Currently, deep learning has emerged as a highly successful technique in machine learning. It has demonstrated the ability to train complex models and extract advanced features from real-world datasets. In deep learning, generative networks can capture intricate relationships between features through deep neural network structures, thereby enhancing the accuracy of predictions.

With the increasing prevalence and integration of various data types, such as genomics, transcriptomics, and tissue pathology data, cancer treatment is shifting towards precision medicine [24]. Utilizing and interpreting multiple high-dimensional data types in translational research or clinical tasks requires a significant amount of time and expertise. This necessitates modeling algorithms capable of learning from a multitude of complex features. Excitingly, deep learning models have the potential to leverage this complexity to provide meaningful insights and identify relevant granular features from diverse data types [25, 26]. Whether it is tabular data or image data, the application of deep learning-based survival analysis models can be achieved by constructing appropriate model architectures and training processes. It is crucial to select suitable model structures based on specific data types and application scenarios, while also performing feature engineering and model optimization to obtain accurate survival predictions. We discuss deep learning-based survival analysis models from two perspectives: utilizing omics-data and image data. In Table 1, we have compiled the datasets used in this study. The WHAS dataset comprises trends in cardiac event occurrences, patient characteristics, and the quality of healthcare in the region. The SUPPORT study aims to understand the communication and decision-making processes among patients, their families, and healthcare providers. It examines medical decisions related to various types of treatment interventions and assesses the impact of these decisions on patient survival and quality of life. METABRIC aims to investigate the classification and molecular characteristics of breast cancer through molecular biology and genetics approaches. TCGA aims

to improve the prevention, diagnosis, and treatment of various cancer types by conducting in-depth research on the genomes and molecular characteristics of cancer. TCGA data provides extensive information related to tumor genomics, mutations, gene expression, methylation, and more. NLST aims to assess the effectiveness of lung CT scans in lung cancer screening.

Table 2. Summary of omics-dataset.

Publication	Type of Disease	Dataset	Variable of input
[27]	Heart attack, Breast(BRCA)	WHAS, SUPPORT, METABRIC	Age, sex, body-mass-index(BMI), left heart failure complications (CHF), and order of MI (MIORD)
[28]	Pan-glioma (LGG/GBM), BRCA, Pan-kidney (KIPAN)	TCGA	Gene expression features, protein expression, copy number, age and stage
[29]	BRCA	METABRIC	Tumor size, number of positive lymph nodes
[30]	Comprehensive	CLINIC	Physiologic variables
[31]	BRCA	TCGA	Gene expression data, miRNA data, Copy number burden (CNB), Tumor mutation burden and clinical information
[32]	BRCA	TCGA	methylation and mRNA data

Omics-Data. Transcriptomics data is currently the most commonly used type of Omics-data. Table 2 provides an overview of the omics data used in related works. Transcriptomic analysis can be employed to assign cancer to clinically meaningful molecular subtypes with diagnostic, prognostic, or therapeutic relevance. Standard computational methods for cancer subtyping, such as SVM or k-Nearest Neighbors (kNN), may be prone to batch effects and can be error-prone, relying only on a small number of feature genes while disregarding important biological information. Deep learning algorithms can overcome these limitations by learning patterns across the entire transcriptome. DeepSurv, based on the semi-parametric Cox proportional hazards model, uses a deep neural network instead of a linear network for prediction [27]. DeepSurv has demonstrated the significant role of deep neural networks in survival analysis tasks. However, this prediction method has limitations when learning from high-dimensional profiles generated from these platforms and relies on expert manual selection of a few features to train the predictive model. SurvivalNet demonstrates how deep learning and Bayesian optimization methods can be applied to predict cancer outcomes, which have been successful in general high-dimensional prediction tasks [28]. SurvivalNet is a Bayesian optimization-based deep survival model that successfully transfers information across diseases to improve prognostic accuracy. However, these models rely on strong parametric assumptions, which are often violated in practice. DeepHit, on the other hand, does not make any assumptions about

underlying stochastic processes and allows for the possibility of time-varying relationships between covariates and risks [29]. Moreover, DeepHit handles competing risks. It introduces a ranking loss in the loss function to address these challenges. DeepHit effectively improves the prediction of correct order pairs by optimizing this loss function:

$$A_{k,i,j} \equiv \mathbf{1}(k^{(i)} = k, s^{(i)} < s^{(j)}) \quad (5)$$

$$L = - \sum_{k=1}^K \alpha_k \sum_{i \neq j} A_{k,i,j} \cdot \eta \left(F_k \left(s^{(i)} | x^{(i)} \right), F_k \left(s^{(i)} | x^{(j)} \right) \right) \quad (6)$$

In this context, where k represents the occurrence of an event, s denotes the time of event occurrence, x represents patient covariates, and $F()$ is the prediction function, $F(s|x)$ represents the estimated risk of an event at time s for a given patient. During the model training process, this function aims to make the model predict patients with longer survival times and lower risk of death as indicated by the labels.

Few studies have considered sequence patterns within the feature space. To address this, (Ren et al., 2019) proposed a deep recursive survival analysis model that utilizes deep learning for fine-grained conditional probability prediction of the data, while incorporating survival analysis to address censoring issues [30]. This approach models the conditional probability of events for each sample, capturing temporal dependencies to predict the likelihood of real event occurrences. It also estimates the survival rates over time for censored data, thereby improving prediction accuracy.

Numerous studies have made further contributions to the richness of input data, allowing the model training to encompass not only single-source data such as gene expression data or biomarkers but also incorporate multiple omics data. SAMMON has been proposed to aggregate and simplify gene expression data and cancer biomarkers for prognostic prediction [31]. Experimental results demonstrate that performance is improved when more omics data is used in model construction. However, incorporating more data into the model also presents challenges. Gene data is characterized by high dimensionality, non-linearity, and sparsity. Directly training the model with these high-dimensional data does not yield optimal results. (Bichindaritz et al., 2021) employed the local maximum quasi-cliques merging (lmQCM) algorithm to reduce the dimensions of mRNA and methylation features and extract clustered feature genes [32]. They introduced an auxiliary ordinal loss on top of the original Cox model to enhance the optimization learning process during training and regularization. The auxiliary loss helps mitigate the problem of vanishing gradients in early layers and aids in reducing the loss of the primary task. Finally, they constructed an ordered Cox risk model for survival analysis and employed the Long Short-Term Memory (LSTM) approach to predict patients' survival risks.

Image Data. Convolutional neural network (CNNs) have been extensively applied in the intersection of computer science and medicine. These end-to-end

Table 3. Summary of image dataset.

Publication	Type of Disease	Dataset	Variable of input
[34]	Lung and brain cancers	NLST, TCGA	Whole Slide Histopathological Images(WSIs)
[35]	Lung cancer	NLST	The screening radiography arm of the NLST
[36]	Lung and colorectal cancers	NLST	WSIs
[37]	BRCA	TCGA	WSIs
[38]	Heart failure	Clinical records	The clinical information, cine MR images, and clinic outcome
[39]	Glioblastoma multiforme (GBM), Lung squamous cell carcinoma (LUSC)	TCGA	WSIs

deep neural networks possess stronger feature extraction capabilities compared to traditional methods [33]. Unlike manual feature selection, CNNs can automatically extract the most discriminative features from images. The Table 3 provides an overview of the imaging data used in the relevant studies. Moreover, they can leverage pre-trained models trained on ImageNet during training, enabling them to be quickly trained and optimized on medical image datasets with limited data, yielding satisfactory results. However, in a high-resolution medical image, only a small portion typically contains the relevant features. The prevalence of negative samples significantly outnumbering positive samples can lead to suboptimal model performance in such scenarios. (Zhu et al., 2017) proposed an effective whole slide image-based survival analysis framework (WSISA) to overcome the aforementioned challenges [34]. To leverage WSIs for survival discrimination, they first extracted hundreds of patches from each WSI through adaptive sampling and then grouped these images into different clusters. Subsequently, they trained an ensemble model to make patient-level predictions based on the cluster-level deep convolutional survival (DeepConvSurv) predictions. This framework efficiently mines and utilizes all discriminative patterns within WSIs to predict patient survival status. (Lu et al., 2019) presented the prediction of patients' mortality risk using X-ray images [35]. Traditional image-based survival prediction models rely on annotated discriminative markers, which limits their scalability to large datasets. When there are no available annotations for the classification task, the multiple instance learning (MIL) framework proves useful for histopathological images. Unlike existing image-based survival models that are restricted to extracting key patches or clusters obtained from whole slide images (WSIs), DeepAttnMISL [36] effectively learns imaging features from WSIs by introducing attention-based MIL pooling and aggregates WSI-level information into patient-level predictions. In current survival models, attention-based aggregation offers more flexibility and adaptability compared to traditional aggregation techniques. (Liu et al., 2022) proposed and experimentally evaluated a multi-resolution deep learning approach for breast cancer survival analysis [37]. This method integrates multiple resolution image data with tumor, lymphocyte, and nucleus segmentation results from deep learning models. The results demon-

strate that this approach significantly improves the performance of deep learning models compared to using only raw image data. However, conventional images capture two-dimensional spatial information as they only contain information within the plane. They cannot capture the temporal trends of patients over time. (Guo et al., 2023) proposed a predictive model based on cardiac images that incorporates temporal sequence information [38]. By fusing the boundary information of cardiac images with the motion field of the heart, they obtained cardiac motion information, which improved the survival probability prediction for heart failure patients. (Li et al., 2023) introduced a patch sampling strategy based on image information entropy and constructed a multi-scale feature fusion network (MSFN) using a self-supervised feature extractor [39]. Specifically, this work employed image information entropy as a criterion to select representative sampling patches, thereby avoiding noise interference caused by random sampling in blank regions. Additionally, a pre-training process was performed on the feature extractor using self-supervised learning mechanisms to enhance feature extraction efficiency. Furthermore, a global-local feature fusion prediction network based on attention mechanisms was constructed to improve the survival prediction of WSIs with comprehensive multi-scale information representation.

Table 4. Summary of multimodal dataset.

Publication	Type of Disease	Dataset	Variable of input
[40]	Lower-Grade Glioma (LGG) and Glioblastoma (GBM)	TCGA	Whole-slide images, clinical and genomic data
[41]	20 cancer types	TCGA	MicroRNAs, gene expression data, and clinical data
[42]	Lung and colorectal cancers	NLST	CT scans, age, gender, and smoking history

Multimodal Data. Multimodal fusion refers to the process of integrating information from two or more modalities to make predictions. In prediction tasks, individual modalities often do not contain all the necessary information required for accurate prediction results. The process of multimodal fusion combines information from multiple modalities to complement each other, expand the coverage of information contained in the input data, enhance the accuracy of prediction results, and improve the robustness of the prediction model (Fig. 1).

Simultaneously combining Omics-data and image data is a key factor in further improving model performance. The Table 4 provides an overview of the multimodal data used in the related studies. (Mobadersany et al., 2018) demonstrated a computational approach that utilizes deep learning to learn patient outcomes from digital pathology images by combining adaptive machine learning algorithms with traditional survival models [40]. This work showcased how survival convolutional neural networks (SCNNs) integrate information from histopathological images and genomic biomarkers into a unified framework to predict time-to-event outcomes. It showed superior prediction accuracy for overall survival

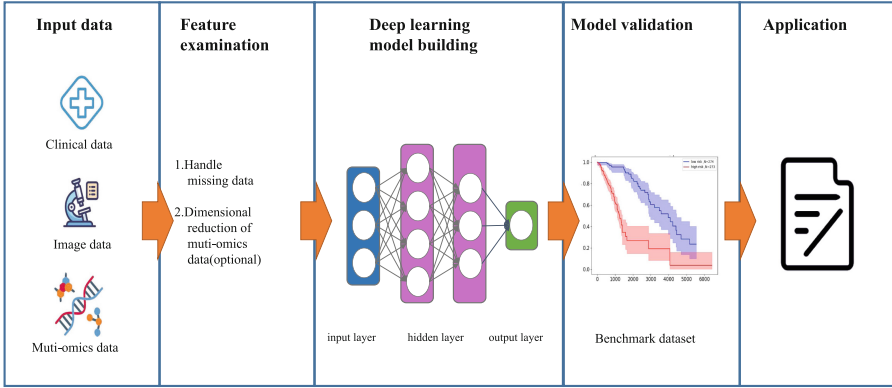


Fig. 1. Workflow for establishing a deep learning model for prognosis prediction.

of glioma patients beyond current clinical models. (Cheerla et al., 2019) further developed a model based on a multimodal neural network that utilizes clinical data, mRNA expression data, microRNA expression data, and WSIs to predict survival of patients across 20 different cancer types [41]. They devised an unsupervised encoder that compresses the four data modalities into a single feature vector per patient, handling missing data using an elastic, multimodal dropout approach. The encoding method is tailored for each data type - using deep highway networks to extract features from clinical and genomic data and employing convolutional neural networks to extract features from WSIs. (Lu et al., 2023) proposed a hybrid CNN-RNN approach to investigate the long-term survival rates of subjects in lung cancer screening studies [42]. This method utilizes a CNN model to capture imaging features from CT scans and employs an RNN model to examine the time series aspect, thereby capturing global information. By combining time series information with multimodal data, the performance of the model is effectively improved.

4 Cluster Based Risk Profile

Clustering is an unsupervised learning method that can uncover hidden patterns and structures within data. Clustering is a valuable tool for data-driven disease discovery and classification. The objective of survival clustering is to map subjects (e.g., users in a social network, patients in medical research) into K clusters ranging from low-risk to high-risk.

4.1 Statistics Based Methods

(Li et al., 2004) proposed an approach to examine survival data by extending the partial least squares (PLS) regression to the framework of the Cox model [43]. They presented a parallel algorithm for constructing latent components.

This algorithm utilized residual iterative least squares fitting and Cox regression fitting to construct predictive components. These components could then be used to build useful survival prediction models and also employed for clustering survival data since the principal components were simultaneously constructed.

(Bair et al., 2004) were the first to explore clustering methods for survival data, introducing a technique known as semi-supervised clustering (SSC) [44]. In their study, they proposed using risk scores from univariate Cox regression as a preselection step to choose variables and then applied k-means clustering to the selected subset of features to discover patient subgroups. In the second part of the method, they employed only clinical data to test the clustering assignments. Using the clinical data, they treated the clustering assignments as the outcome variable and applied a classification algorithm. The classification algorithm performed well, indicating correct identification of the clustering assignments. However, both the regression and survival models utilized principal components. Since principal components may not capture a large portion of the variance present in the data, these methods cannot guarantee that these components are relevant to the outcomes. Therefore, (Bair et al. 2006) proposed a semi-supervised approach called supervised principal components (SPC) [45]. In this method, they computed univariate Cox scores for each feature and selected the most important features by choosing only those with the best Cox scores.

Sparse clustering methods and semi-supervised clustering approaches rely on the number of features that have been characterized as “salient,” and therefore, these methods have notable limitations. (Gaynor et al., 2017) proposed an enhanced method called pre-weighted sparse clustering to overcome the limitations of sparse clustering [46]. Its objective is to overcome the limitations of traditional sparse clustering by identifying features that have different means within clusters. This approach can identify features that exhibit variations in their average values across clusters.

4.2 Traditional Machine Learning Based Methods

(Zhang et al. 2016) employed a hybrid approach consisting of statistical and machine learning methods, focusing on cluster discovery in clinical and gene expression data [47]. They utilized penalized logistic regression and penalized proportional hazards models along with an expectation-maximization algorithm to select only the most relevant clinical features associated with the event of interest. This approach allowed them to identify the most important clinical features specifically related to the event of interest. (Mouli et al., 2017) proposed a decision tree-based method aimed at achieving survival clustering [48]. The ultimate goal was to cluster examination data and identify two or more populations with different risk levels. The objective was to determine clusters where the survival distributions differed. (Ahlqvist et al., 2018) utilized Cox regression to explore differences between subgroups of diabetes patients discovered through k-means and hierarchical clustering [49]. This method divided patients into five subgroups based on distinct disease progression and risk of diabetes complications. This novel subtyping could potentially contribute to personalized and

targeted early treatments for patients who would benefit the most, representing an initial step towards precision medicine in diabetes.

Table 5. Summary of cluster dataset.

Publication	Type of Disease	Dataset	Variable of input
[50]	Nondisease	Friendster	age, gender, marital status, occupation, and interest
[51]	Acute coronary syndrome	Private	disease types, demographics, personal disease history, comorbidities, habits, laboratory test results, and procedures
[52]	Type2 diabetes, heart dise	FLCHAIN, SUPPORT, SEER, EHR, SLEEP, FRAMINGHAM	serum data, breast cancer subcohort and clinical data
[53]	BRCA	SUPPORT, METABRIC, SEER	Age, gender, race, education, income, physiological measurements, co-morbidity information, gene expressions, and clinical variables
[54]	Alzheimer’s disease(AD)	EHR	demographic information, laboratory tests, and diagnoses and symptoms
[55]	Cardiovascular disease	ALLHAT-A, ALLHAT-B, ACCORD	Clinical information on the use of different concentrations of drugs
[56]	22 high-grade glioma patients, non-small cell	SUPPORT, HGG, Hemodialysis, FLChain, NSCL	Observational cohort and CT image

4.3 Neural Network Based Methods

Deep neural networks play a significant role in the advancement of clustering tasks. They are commonly used to handle large-scale and high-dimensional data, such as images, speech, and textual data, in the field of computational biology. The Table 5 provides an overview of the data used in deep learning clustering research. Traditional survival methods assume the existence of explicit end-of-life signals or introduce them artificially using predefined timeouts (Mouli et al., 2018). They proposed a deep clustering approach that distinguishes long-term and short-term survivors based on a modified Kuiper statistic, even in the absence of end-of-life signals [50]. In their study, they introduced a loss function that utilizes an enhanced Kuiper statistic to differentiate the empirical survival distributions of clusters. By optimizing this loss, a deep neural network is learned to softly cluster users into survival groups. (Xia et al., 2019) employed a multi-task learning approach for outcome-driven clustering of patients with acute coronary syndrome [51]. The proposed method utilized an attention-based

multi-task neural network as the modeling framework, which includes patient state learning, cluster analysis, and feature importance analysis.

However, traditional survival analysis methods estimate risk scores or personalized event time distributions that depend on covariates. In practice, due to (unknown) subpopulations having different risk profiles or survival distributions, there often exists substantial population-level phenotypic heterogeneity. Therefore, in survival analysis, there is an unmet need to identify subgroups with distinct risk profiles while simultaneously considering accurate personalized event time predictions. Methods addressing this need may improve the characterization of individual outcomes by leveraging the regularities within subpopulations, thereby accounting for population-level heterogeneity. (Chapfuwa et al., 2020) proposed a Bayesian nonparametric method that represents observations (subjects) in a clustering latent space and encourages accurate time-to-event predictions and clustering (subpopulations) with distinct risk profiles [52]. (Nagpal et al., 2021) have explored similar techniques, introducing a finite mixture of Weibull distributions known as the Deep Survival Machine (DSM) [53]. DSM fits a survival regression model mixture on the representations learned by an encoder neural network. From a modeling perspective, the aforementioned methods focus on outcome-driven clustering, where they fully recover clusters with distinct survival distribution characteristics. In this work, the Deep Cox Mixture (DCM) is introduced, which jointly fits a VAE and Cox regression mixture without specifying a generative model. The loss of DCM is derived by combining the VAE loss and likelihood estimation of survival time.

Previous research has primarily utilized imaging or cognitive data, with limitations in data breadth and sample size. Data-driven models have not been able to perform well in these cases. Certain diseases exhibit a high degree of heterogeneity, such as Alzheimer’s disease (AD), where different trajectories and outcomes are observed in clinical populations. (Alexander et al., 2021) identified AD patients using a previously validated rule-based phenotype algorithm from the Clinical Practice Research Datalink (CPRD), which contains primary care electronic health records [54]. They extracted and incorporated a range of comorbidities, symptoms, and demographic features as patient characteristics, thus expanding the breadth of data. However, this approach did not consider the evaluation of treatment effects concerning clinical interventions involving continuous time-to-event outcomes, such as time to death, readmission, or composite events that may be subject to review. In such cases, counterfactual inference is required to disentangle the effects of confounding physiological features affecting baseline survival rates from the effects of the interventions being evaluated. (Nagpal et al., 2022) proposed a latent variable approach to simulate heterogeneous treatment effects, suggesting that an individual can belong to one of several latent clusters with different response characteristics [55]. Experimental results demonstrate that this latent structure can modulate baseline survival rates and help determine the effects of interventions. However, clustering of survival data remains an underexplored problem. In this scenario, only a few methods have been proposed, either with limited scalability in high-dimensional unstructured

data or focused on discovering purely outcome-driven clusters, i.e., clusters solely based on survival time as the defining feature. The latter may fail in applications where individual survival distribution information alone is insufficient for stratifying the population. For example, patient groups with similar survival outcomes may exhibit vastly different responses to the same treatment. To address these challenges, (Manduchiy et al., 2022) introduced a novel survival data clustering approach called Variational Deep Survival Clustering (VaDeSC), which discovers patient groups with distinct characteristics in terms of the underlying mechanisms generating survival outcomes [56]. It extends previous variational methods used for unsupervised deep clustering by incorporating survival models specific to each cluster within the generative process. VaDeSC focuses not only on survival but also captures the heterogeneity in the relationship between covariates and survival outcomes.

5 Conclusions and Future Directions

Despite the numerous scientific reports on the application of machine learning techniques for time-to-event prediction, there has been relatively less research on survival clustering techniques. Survival clustering techniques are particularly useful when there is a need to identify unknown subpopulations within an entire dataset. They can discover clusters with significantly different survival capabilities, which cannot be achieved by traditional clustering techniques. These techniques focus on finding clusters with distinct survival distributions, providing a unique perspective to understand the characteristics of the dataset.

Many research works in the field of survival analysis are currently exploring the application of deep learning methods, which have powerful modeling capabilities and predictive performance. However, a major limitation of deep learning models is their lack of interpretability. This means that although deep learning models can generate accurate prediction results, it is often challenging to explain how the model arrives at those predictions and what features and patterns it relies on for decision-making. Further more, the research efforts in exploring key biomarkers are also relatively scarce in the field.

Enhancing Model Interpretability: One significant challenge in the application of machine learning models to survival analysis lies in their inherent complexity and the difficulty in understanding how they arrive at specific predictions. As the use of machine learning in this domain continues to grow, there is a pressing need to develop interpretable models that can provide meaningful insights into the underlying biological mechanisms governing survival outcomes. Future research efforts should focus on incorporating techniques such as feature importance analysis, attention mechanisms, and visualizations to shed light on the decision-making processes of these models. By achieving better interpretability, researchers can gain a deeper understanding of the relationships between genomic, clinical, and imaging features, ultimately leading to more reliable and clinically actionable predictions.

Identifying Novel Biomarkers: While machine learning techniques have shown great promise in survival analysis, there remains untapped potential in the discovery of novel and robust biomarkers that can accurately predict patient outcomes. Leveraging multi-omics data integration and advanced feature selection methods can facilitate the identification of previously unrecognized biomarkers with strong prognostic significance. Additionally, collaborative efforts between computational biologists, bioinformaticians, and domain experts can drive the development of innovative approaches to uncover hidden patterns and relationships within complex biological datasets. The integration of these novel biomarkers into clinical practice has the potential to revolutionize patient risk stratification, enabling tailored treatment strategies and personalized medicine.

By addressing these challenges, it not only contributes to enhancing the feasibility and acceptability of survival analysis methods in medicine and other fields but also promotes the advancement of survival analysis techniques to meet broader application needs. Furthermore, it provides more reliable decision support for clinical practice and disease management.

References

1. Guo, S.: *Survival Analysis*. Oxford University Press (2010)
2. Emmert-Streib, F., Dehmer, M.: Introduction to survival analysis in practice. *Mach. Learn. Knowl. Extr.* **1**(3), 1013–1038 (2019)
3. Nezhad, M.Z., Sadati, N., Yang, K., et al.: A deep active survival analysis approach for precision treatment recommendations: application of prostate cancer. *Expert Syst. Appl.* **115**, 16–26 (2019)
4. Singer, J.D., Willett, J.B.: Modeling the days of our lives: using survival analysis when designing and analyzing longitudinal studies of duration and the timing of events. *Psychol. Bull.* **110**(2), 268 (1991)
5. Klein, J.P., Moeschberger, M.L.: *Survival Analysis*. SBH, Springer, New York (2003). <https://doi.org/10.1007/b97377>
6. Lee, E.T., Wang, J.: *Statistical Methods for Survival Data Analysis*. John Wiley & Sons (2003)
7. Marubini, E., Valsecchi, M.G.: *Analysing Survival Data from Clinical Trials and Observational Studies*. John Wiley & Sons (2004)
8. Kaplan, E.L., Meier, P.: Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**(282), 457–481 (1958)
9. Andersen, P.K., Borgan, O., Gill, R.D., et al.: *Statistical Models Based on Counting Processes*. Springer Science & Business Media (2012)
10. Cutler, S.J., Ederer, F.: Maximum utilization of the life table method in analyzing survival. *J. Chronic Dis.* **8**(6), 699–712 (1958)
11. Lin, D.Y., Wei, L.J., Ying, Z.: Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**(3), 557–572 (1993)
12. Bou-Hamad, I., Larocque, D., Ben-Ameur, H.: A review of survival trees (2011)
13. Ciampi, A., Chang, C.H., Hogg, S., McKinney, S.: Recursive partition: a versatile method for exploratory-data analysis in biostatistics. In: MacNeill, I.B., Umphrey, G.J., Donner, A., Jandhyala, V.K. (eds.) *Biostatistics*. The University of Western Ontario Series in Philosophy of Science, vol. 38. Springer, Dordrecht (1987). https://doi.org/10.1007/978-94-009-4794-8_2

14. Ishwaran, H., Kogalur, U.B., Blackstone, E.H., et al.: Random survival forests (2008)
15. Shahraki, H.R., Salehi, A., Zare, N.: Survival prognostic factors of male breast cancer in Southern Iran: a LASSO-Cox regression approach. *Asian Pac. J. Cancer Prev.* **16**(15), 6773–6777 (2015)
16. Khan, F.M., Zubek, V.B.: Support vector regression for censored data (SVRc): a novel tool for survival analysis. In: 2008 Eighth IEEE International Conference on Data Mining, pp. 863–868. IEEE (2008)
17. Kononenko, I.: Inductive and Bayesian learning in medical diagnosis. *Appl. Artif. Intell. Int. J.* **7**(4), 317–337 (1993)
18. Ibrahim, J.G., Chen, M.H., Sinha, D., et al.: Bayesian survival analysis. Springer, New York (2001). <https://doi.org/10.1007/978-1-4757-3447-8>
19. Fard, M.J., Wang, P., Chawla, S., et al.: A Bayesian perspective on early stage event prediction in longitudinal data. *IEEE Trans. Knowl. Data Eng.* **28**(12), 3126–3139 (2016)
20. Van Belle, V., Pelckmans, K., Suykens, J.A.K., et al.: Support vector machines for survival analysis. In: Proceedings of the Third International Conference on Computational Intelligence in Medicine and Healthcare (CIMED2007), pp. 1–8 (2007)
21. Shivaswamy, P.K., Chu, W., Jansche, M.: A support vector approach to censored targets. In: Seventh IEEE International Conference on Data Mining (ICDM 2007), pp. 655–660. IEEE (2007)
22. Van Belle, V., Pelckmans, K., Suykens, J.A.K., et al.: Support vector machines for survival analysis. In: Proceedings of the Third International Conference on Computational Intelligence in Medicine and Healthcare (CIMED 2007), pp. 1–8 (2007)
23. Van Belle, V., Pelckmans, K., Van Huffel, S., et al.: Support vector methods for survival analysis: a comparison between ranking and regression approaches. *Artif. Intell. Med.* **53**(2), 107–118 (2011)
24. Tran, K.A., Kondrashova, O., Bradley, A., et al.: Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med.* **13**(1), 1–17 (2021)
25. Wainberg, M., Merico, D., DeLong, A., et al.: Deep learning in biomedicine. *Nat. Biotechnol.* **36**(9), 829–838 (2018)
26. Zou, J., Huss, M., Abid, A., et al.: A primer on deep learning in genomics. *Nat. Genet.* **51**(1), 12–18 (2019)
27. Katzman, J.L., Shaham, U., Cloninger, A., et al.: DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med. Res. Methodol.* **18**(1), 1–12 (2018)
28. Yousefi, S., Amrollahi, F., Amgad, M., et al.: Predicting clinical outcomes from large scale cancer genomic profiles with deep survival models. *Sci. Rep.* **7**(1), 11707 (2017)
29. Lee, C., Zame, W., Yoon, J., et al.: DeepHit: a deep learning approach to survival analysis with competing risks. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1 (2018)
30. Ren, K., Qin, J., Zheng, L., et al.: Deep recurrent survival analysis. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 33, no. 01, pp. 4798–4805 (2019)
31. Huang, Z., Zhan, X., Xiang, S., et al.: SALMON: survival analysis learning with multi-omics neural networks on breast cancer. *Front. Genet.* **10**, 166 (2019)

32. Bichindaritz, I., Liu, G., Bartlett, C.: Integrative survival analysis of breast cancer with gene expression and DNA methylation data. *Bioinformatics* **37**(17), 2601–2608 (2021)
33. Raza, R., Zulfiqar, F., Tariq, S., et al.: Melanoma classification from Dermoscopy images using ensemble of convolutional neural networks. *Mathematics* **10**(1), 26 (2022)
34. Zhu, X., Yao, J., Zhu, F., et al.: 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)
35. Lu, M.T., Ivanov, A., Mayrhofer, T., et al.: Deep learning to assess long-term mortality from chest radiographs. *JAMA Netw. Open* **2**(7), e197416–e197416 (2019)
36. Yao, J., Zhu, X., Jonnagaddala, J., et al.: Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks. *Med. Image Anal.* **65**, 101789 (2020)
37. Liu, H., Kurc, T.: Deep learning for survival analysis in breast cancer with whole slide image data. *Bioinformatics* **38**(14), 3629–3637 (2022)
38. Guo, S., Zhang, H., Gao, Y., et al.: Survival prediction of heart failure patients using motion-based analysis method. *Comput. Methods Programs Biomed.* **236**, 107547 (2023)
39. Li, L., Liang, Y., Shao, M., et al.: Self-supervised learning-based Multi-Scale feature Fusion Network for survival analysis from whole slide images. *Comput. Biol. Med.* **153**, 106482 (2023)
40. Mobadersany, P., Yousefi, S., Amgad, M., et al.: Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc. Natl. Acad. Sci.* **115**(13), E2970–E2979 (2018)
41. Cheerla, A., Gevaert, O.: Deep learning with multimodal representation for pancreatic cancer prognosis prediction. *Bioinformatics* **35**(14), i446–i454 (2019)
42. Lu, Y., Aslani, S., Zhao, A., et al.: A hybrid CNN-RNN approach for survival analysis in a Lung Cancer Screening study (2023). arXiv preprint [arXiv:2303.10789](https://arxiv.org/abs/2303.10789)
43. Li, H., Gui, J.: Partial Cox regression analysis for high-dimensional microarray gene expression data. *Bioinformatics* **20**(suppl_1), i208–i215 (2004)
44. Bair, E., Tibshirani, R.: Semi-supervised methods to predict patient survival from gene expression data. *PLoS Biol.* **2**(4), e108 (2004)
45. Bair, E., Hastie, T., Paul, D., et al.: Prediction by supervised principal components. *J. Am. Stat. Assoc.* **101**(473), 119–137 (2006)
46. Gaynor, S., Bair, E.: Identification of relevant subtypes via preweighted sparse clustering. *Comput. Stat. Data Anal.* **116**, 139–154 (2017)
47. Zhang, W., Tang, J., Wang, N.: Using the machine learning approach to predict patient survival from high-dimensional survival data. In: *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1234–1238. IEEE (2016)
48. Mouli, S.C., Naik, A., Ribeiro, B., et al.: Identifying user survival types via clustering of censored social network data (2017). arXiv preprint [arXiv:1703.03401](https://arxiv.org/abs/1703.03401)
49. Ahlqvist, E., Storm, P., Käräjämäki, A., et al.: Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diab. Endocrinol.* **6**(5), 361–369 (2018)
50. Mouli, S.C., Ribeiro, B., Neville, J.: A Deep Learning Approach for Survival Clustering without End-of-life Signals (2018)
51. Xia, E., Du, X., Mei, J., et al.: Outcome-driven clustering of acute coronary syndrome patients using multi-task neural network with attention. In: *MedInfo*, pp. 457–461 (2019)

52. Chapfuwa, P., Li, C., Mehta, N., et al.: Survival cluster analysis. In: Proceedings of the ACM Conference on Health, Inference, and Learning, pp. 60–68 (2020)
53. Nagpal, C., Li, X., Dubrawski, A.: Deep survival machines: fully parametric survival regression and representation learning for censored data with competing risks. *IEEE J. Biomed. Health Inform.* **25**(8), 3163–3175 (2021)
54. Alexander, N., Alexander, D.C., Barkhof, F., et al.: Identifying and evaluating clinical subtypes of Alzheimer’s disease in care electronic health records using unsupervised machine learning. *BMC Med. Inform. Decis. Mak.* **21**(1), 1–13 (2021)
55. Nagpal, C., Goswami, M., Dufendach, K., et al.: Counterfactual phenotyping with censored time-to-events. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 3634–3644 (2022)
56. Manduchi, L., Marcinkevičs, R., Massi, M.C., et al.: A deep variational approach to clustering survival data (2021). arXiv preprint [arXiv:2106.05763](https://arxiv.org/abs/2106.05763)