A Deep Active Survival Analysis approach for precision treatment recommendations: Application of prostate cancer

Milad Zafar Nezhad [a,*], Najibesadat Sadati [a], Kai Yang [a], Dongxiao Zhu [b]

[a] Department of Industrial and Systems Engineering, Wayne State University, 4815 Fourth Street, Detroit, MI 48202, USA
[b] Department of Computer Science, Wayne State University, 5057 Woodward Ave, Detroit, MI 48202, USA

ARTICLE INFO

ABSTRACT

Survival analysis has been developed and applied in the number of areas including manufacturing, finance, economics and healthcare. In healthcare domain, usually clinical data are high-dimensional, sparse and complex and sometimes there exists few amount of time-to-event (labeled) instances. Therefore building an accurate survival model from electronic health records is challenging. With this motivation, we address this issue and provide a new survival analysis framework using deep learning and active learning with a novel sampling strategy. First, our approach provides better representation with lower dimensions from clinical features using labeled (time-to-event) and unlabeled (censored) instances and then actively trains the survival model by labeling the censored data using an oracle. As a clinical assistive tool, we introduce a simple effective treatment recommendation approach based on our survival model. In the experimental study, we apply our approach on SEER-Medicare data related to prostate cancer among African–Americans and white patients. The results indicate that our approach outperforms significantly than baseline models.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction

Survival analysis has been applied in several real-world applications such as healthcare, manufacturing and engineering in order to model time until the occurrence of a future event of interest (e.g., biological death or mechanical failure) (Hosmer, Lemeshow, & May, 2011). Censoring attribute of survival data makes survival analysis different from the other prediction approaches. One popular survival model is the Cox Proportional Hazards model (CPH) (Cox, 1992) which models the risk of an event happening based on linear combination of the covariates (risk factors). The major problem of Cox-based models is linear relationship assumption between covariates and the time of event occurrence. Hence, there have been developed several models to handle non-linear relationship in survival analysis like as survival neural network and survival random forest models (Ishwaran et al., 2014).

In the healthcare area, medical researchers applied survival analysis on Electronic Health Records (EHRs) to evaluate the significance of many risk factors in outcomes such as survival rates or cancer recurrence and subsequently recommend treatment schemes. There exist two specific challenges in survival analysis from EHRs: (1) Clinical data is usually high dimensional, sparse and time-dependent which in this case applying traditional survival approaches do not perform well enough to estimate the risk of a medical event, (2) In many health survival applications, labeled data (time-to-event instances) are small, time-consuming and expensive to collect. In this situation, it is hard to learn a survival model based on traditional approaches which able to predict the relative risk of patients precisely.

To address the first challenge, recently, semi-supervised learning using deep feature representation has been applied in number of areas and could improve the performance of different machine learning tasks as well as survival analysis. In other words, applying unsupervised learning using deep learning can reduce the complexity of raw data and provide robust features with lower dimensions (LeCun, Bengio, & Hinton, 2015). Using these represented features in the supervised learning algorithms (e.g., survival models) establishes a semi-supervised learning framework which achieves higher performance.

To overcome the second challenge, active learning is well suited to get high accuracy when the labeled instances are small or labeling is expensive and time-consuming (Settles, 2010). Active learning approach from censored data has been rarely addressed in the literature. However it has been widely used in the other aspects of health informatics where the labeled data are scarce.

* Corresponding author.
E-mail addresses: m.zafarnezhad@wayne.edu (M.Z. Nezhad), n.sadati@wayne.edu (N. Sadati), kai.yang@wayne.edu (K. Yang), dzhu@wayne.edu (D. Zhu).

According to the current works in the literature, no research has been conducted to overcome these two challenges by developing an integrated method for survival analysis. Although, there exist few studies in the literature that focus to one of these challenges, our goal is to address both of them simultaneously. In addition, several applications (especially in the healthcare domain) demand to develop such that integrated approach when they deal with a few amount of labeled instances that are high-dimensional and training a precise survival analysis model is difficult based on the current baselines. To address this research gap, first, we propose a novel survival analysis approach using deep learning and active learning termed DASA. Our method is capable to learn more accurate survival model using high dimensional and small size EHRs in comparison with some baseline survival approaches. Second, we introduce a personalized treatment recommendation approach based on our survival analysis model which can compare the relative risks (or survival times) associate with different treatment plans and assign the better one. We evaluate our approach using SEER-Medicare dataset related to prostate cancer. We consider two racial subgroup of patients (African-American and whites) in our analysis and apply our model on each dataset separately.

Our contributions in this research lie into three folds: (1) To the best of our knowledge, we propose the first Deep Active Survival Analysis approach with promising performance, (2) In our active learning framework we develop a new sampling strategy specifically for survival analysis and (3) Our model with proposed treatment recommendation approach has highly potential to apply for evaluation of new treatment effect on new patients where the labeled data is scarce.

2. Background

In this section, we review some basic concepts for modeling of survival analysis, active learning and deep learning.

2.1. Introduction to survival analysis

Survival analysis is a kind of statistical modeling where the main goal is to analyze and model time until the occurrence of an event of interest. The challenging characteristic of survival data is the fact that time-to-event of interest for many instances is unknown because the event might not have happened during the period of study or missing tracking occurred caused by other events. This concept is called censoring which makes the survival analysis different (Wang, Li, & Reddy, 2017). The special case of censoring is when the observed survival time is less than or equals to the true event time called right-censoring, the main focus of our study.

Since the censored data is present in survival analysis, the standard statistical and machine learning approaches are not appropriate to analyze and predict time-to-event outcome because those approaches miss the censored/right-censored instances. Survival modeling provides different statistical approaches to analyze such censored data in many real-world applications.

In survival analysis, a given instance i, represented by a triplet (X_i, δ_i, T_i) where X_i refers to the instance characteristics and Ti indicates time-to-event of the instance. If the event of interest is observed, Ti corresponds to the time between baseline time and the time of event happening, in this case $\delta_i = 1$. If the instance event is not observed and its time to event is greater than the observation time, Ti corresponds to the time between baseline time and end of the observation, and the event indicator is $\delta_i = 0$. The goal of survival analysis is to estimate the time to the event of interest (T) for a new instance X_j (Wang et al., 2017).

2.1.1. Survival and hazard functions

Survival and hazard functions are the two main functions in survival modeling. The survival function indicates the probability that the time to the event of interest is not less than a determined time (t) (Kleinbaum & Klein, 2010). This function (S) is denoted by following formula:

$$S(t) = Pr(T > t) \tag{1}$$

The initial value of survival function is 1 when $t = 0$ and it monotonically decreases with t. The second function, hazard function indicates the rate of occurrence of the event at time t given that no event occurred earlier. It describes the risk of failure (dying) changing over time. The hazard function (or hazard rate or failure rate) is defined as following (Kleinbaum & Klein, 2010):

$$h(t) = \lim_{\delta(t) \to 0} \frac{Pr(t \leq T \leq t + \delta(t) | T \geq t)}{\delta(t)} \tag{2}$$

Survival and hazard function are non-negative functions. While the survival function decreases over time, the shape of a hazard function can be in different forms: increasing, decreasing, constant, or U-shaped.

2.1.2. Cox Proportional Hazards (CPH) model

There exist several models for survival analysis in the literature. Among all, Cox Proportional Hazards (CPH) model (Cox, 1992) is the most popular model for survival analysis. CPH estimates the hazard function $h(x)$ as a regression formulation:

$$h(t, X_i) = h_0 \, exp(X_i \beta) \tag{3}$$

where h_0 is the baseline hazard function which can be an arbitrary nonnegative function of time and X_i refers to covariate vector for instance i, and β is the coefficient vector estimated after survival model training by maximizing the cox partial likelihood. Because the baseline hazard function $h_0(t)$ in CPH is not determined, we cannot use the standard likelihood function in training process. The partial likelihood is the product of the probability of each instance i at event time T_i that the event has happened for that instance, over the summation of instances (R_j) probability who are still at risk in this time (T_i) (Cox, 1992):

$$L(\beta) = \prod_{i=, \delta_i=1} \frac{exp(X_i \beta)}{\sum_{j \in R_j} exp(X_j \beta)} \tag{4}$$

2.1.3. Evaluation metric for survival analysis

Since the censored instances exist in survival data, the standard evaluation metrics such as mean squared error and R-squared are not appropriate for evaluating the performance of survival analysis (Heagerty & Zheng, 2005). In survival analysis, the most popular evaluation metric is based on the relative risk of an event for different instances called concordance index or c-index. This measure is defined as following formula:

$$\frac{1}{N} \sum_{i, \delta_i=1} \sum_{j, y_i < y_j} I[S(\hat{y}_i | X_i) < S(\hat{y}_j | X_j)] \tag{5}$$

Where N refers to the all comparable instance pairs and S is the survival function. The main motivation for using c-index in survival analysis is originated from the fact that the medical doctors and researchers are often more interested in measuring the relative risk of a disease among patients with different risk factors, than the survival times of patients.

In general, the survival analysis models can be divided into two main categories: (1) statistical methods including non-parametric, semi-parametric and parametric and (2) machine learning based methods such as survival trees, Bayesian methods, neural networks and random survival forests. Readers for more comprehensive review can refer to the recent review provided by Wang et al. (2017).
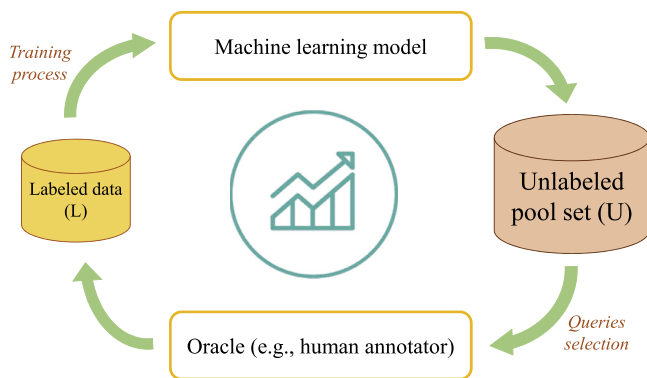
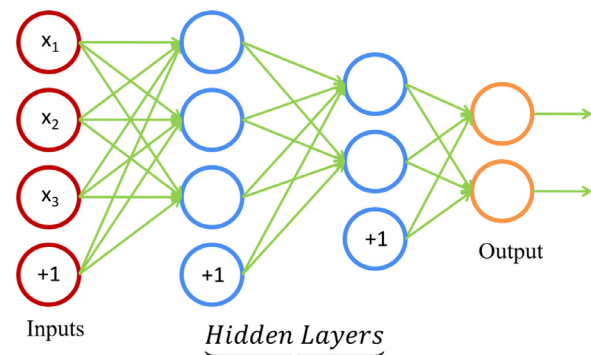**Fig. 1.** The pool-based active learning approach (Settles, 2010).

Fig. 2. Multi layers neural network (deep network).

2.2. Introduction to active learning

Active learning is a subfield of machine learning and statistical modeling. The goal of an active learner is the same as a passive learner but the key idea behind active learning is that a machine learning algorithm can lead to better performance with fewer training labels if it can select the data for learning. An active learner chooses queries, usually in the form of unlabeled data instances to be labeled by an oracle which can be a human annotator, human expert and prior knowledge. Active learning is very efficient in many data-driven applications, where there exist numerous unlabeled data but labels are rare, time-consuming, or expensive to be labeled (Settles, 2010).

Since large amounts of unlabeled data is nowadays often available and can be easily collected by automatic processes, active learning would be demanding in modern applications in order to reduce the cost of labeling. The active learning framework overcomes the challenge of insufficient labeled data by efficiently modeling the process of obtaining labels for unlabeled data. The advantage is that the active learner just requires to query the labels of a few, carefully selected instances during the iterative process in order to achieve more accurate learner (Hsu, 2010).

There exist several approaches/scenarios in which active learners ask queries. The three main approaches widely used in the literature are (Settles, 2010): (1) membership query synthesis (Angluin, 1988), (2) stream-based selective sampling (Atlas, Cohn, & Ladner, 1990), and 3) pool-based sampling (Lewis & Gale, 1994). For all approaches, there are also several different query strategies that have been developed to decide which unlabeled instances should be selected. Among above three approaches, pool-based sampling is most popular in many real-world applications. This approach has been demonstrated in Fig. 1:

According to Fig. 1, in pool-based sampling approach, a learner may start to be trained with a few number of labeled instances (L), then request labels for one or more carefully selected unlabeled instances (U) using an oracle. After labeling, the new instance is simply added to the labeled set(L), and the learner proceeds training process in a standard supervised way. This process continues up to a specified number of iterations or to achieve desired performance.

2.3. Introduction to deep learning

In a simple definition, deep learning or deep machine learning refers to use of a neural network with multiple layers of hidden nodes between input and output where the deep architectures are constructed by several levels of non-linear operations (Fig. 2).

According to Fig. 2, deep network has the same architecture like as traditional neural network with higher number of hidden layers. The main difference between deep network and traditional neural network is the algorithms developed for training deep architecture which are faster and lead to stronger results (Bengio, Courville, & Vincent, 2013).

Deep Learning is including representation learning algorithms that transforms raw features to higher-level abstraction by using a deep network composed several hidden layers (Bengio, 2009). In another word, deep learning applies computational approaches, which have multiple non-linear transformations to train data representation through several levels of abstraction (LeCun et al., 2015; Nezhad, Zhu, Sadati, & Yang, 2018).

Deep learning applications include different areas. The most popular ones are speech detection, image recognition, automatic text generation and health informatics (LeCun et al., 2015). In healthcare domain with explosive increase of large and high-dimensional datasets, deep learning with great performance outperformed some traditional methods in medical features representation and it showed strong potential for feature engineering and dimensionality reduction (Mamoshina, Vieira, Putin, & Zhavoronkov, 2016).

Readers for more detail about applications of deep learning in health informatics can refer to recent review papers provided by Miotto, Wang, Wang, Jiang, and Dudley (2017); Shickel, Tighe, Bihorac, and Rashidi (2017), and Ravi et al. (2017).

3. Related works

Deep learning and active learning as two advanced machine learning methods have been applied in different areas but there exist a few research in the literature that use the benefit of deep learning or active learning in survival analysis. In this section we review studies in three different categories: (1) Methods proposed by using active learning for survival analysis, (2) Studies used deep learning to develop new survival analysis models and (3) Methods proposed based on deep learning and active learning but are not designed for survival analysis.

In the first category, Vinzamuri, Li, and Reddy (2014) provided the first ever active learning framework for survival analysis (this study is the only work in this domain). Authors proposed a novel sampling strategy based on discriminative gradient for selecting the best candidate from the unlabeled pool set. Finally, they evaluated their model performance using public EHRs datasets and compared it with some state of the art survival regression methods. Although their approach demonstrated a good performance, it is only developed based on regularized Cox regression survival model which is limited to linear relationship assumption between covariates and survival time. In the other hand, the authors did

not evaluate the performance of their proposed model using high-dimensional data.

In the deep learning domain (second category), there exist few studies which developed survival analysis framework using deep learning recently. Ranganath, Perotte, Elhadad, and Blei (2016) proposed a new survival model using deep learning termed deep survival analysis. They used Deep Exponential Family (DEF) for capturing complex dependencies from clinical features including laboratory measurements, diagnosis, and medications codes. They applied their model on a large EHR dataset related to coronary heart disease. In the other research (Luck, Sylvain, Cardinal, Lodi, & Bengio, 2017), authors introduced a new deep learning approach which can directly predict the survival times for graft patients using foundations of multi-task learning. They demonstrated that their model outperforms usual survival analysis models such as cox proportional hazard model in terms of prediction quality and concordance index.

Katzman et al. (2018) proposed a Cox Proportional Hazards deep multi-layer perceptron called DeepSurv to predict risk of event occurrence for patient and provided personalized treatment recommendations. They performed their approach on simulated and real-world datasets for testing and evaluation. Finally, they used DeepSurv on real medical studies to illustrate how it can provide treatment recommendations. In the other research, (Lee, Zame, Yoon, & van der Schaar, 2018) introduced a different approach called DeepHit which employs deep architecture to estimate the survival times distribution. They used neural network including two types of sub-networks: (1) a single shared sub-network and (2) family of cause-specific sub-networks. They evaluated their method based on real and synthetic datasets which illustrate that DeepHit leads to better performance in comparison with state of the art methods.

Although these survival models developed by using deep learning are well suited for high-dimensional survival data, they are not the best choice when labeled instances are scarce, it seems more efforts should be accomplished to improve them in such situations. The other drawback of these deep learning based survival model is related to interpretability. Most of research discussed above did not provide interpretable framework for treatment recommendation or survival risk analysis while their proposed method is based on deep representation of original features (risk factors) through multiple non-linear transformations.

There are a few studies that develop deep active learning methods for some machine learning tasks (third category). For example, Zhou, Chen, and Wang (2013) developed a semi-supervised learning framework termed active deep network (ADN) for sentiment analysis. They used restricted Boltzmann machines (RBM) for feature learning based on labeled reviews and large amount of unlabeled reviews, then applied gradient-descent based supervised learning for fine tuning and constructing semi-supervised framework. Finally they used active learning in their framework to improve model performance. In the other study, Liu, Zhang, and Eom (2017) proposed a deep active learning approach using Deep Belief Network (DBN) for classifying hyperspectral images in remote sensing application. All of these studies are appropriate for applying on high-dimensional data when labeled instances are rare (our focus in this research) but they are not developed for survival analysis. In other words, since there exist censored instances in the survival data, the deep active learning design should be different and it is necessary to introduce new approach for survival data specifically.

A summary of our review has been illustrated in Table 1 which indicates there exist no study to develop a survival analysis approach using both deep learning and active learning. We address this gap in the literature to propose a deep active learning framework for survival analysis.

4. Methodology

The method developed in this research is an active learning based survival analysis using a novel sampling strategy. In our model, we apply deep learning for feature reduction and extraction, when data is high-dimensional, complex and sparse. Since in survival analysis we deal with censored and uncensored instances, the active learning design should be different from the regular approach. In our framework, we consider censored and uncensored instances in the training set as survival analysis needs both instances in the training process and we consider uncensored data as unlabeled instances in the pool set because their labels (time to event) are unknown.

The general framework in our survival analysis approach includes two main steps: (1) Deep feature learning for survival data and (2) Active learning based survival analysis. Since deep learning showed a great performance in the feature representation of medical data in different supervised and unsupervised machine learning tasks, in the first step we do unsupervised learning using deep learning to represent features in higher level abstractions and extract data into lower dimensions. Among different types of deep networks in the literature, four deep architectures are more popular in the health domain (Mamoshina et al., 2016) including: 1- Convolutional Neural Network (CNN), 2- Restricted Boltzmann Machine (RBM), 3- Deep Belief Network (DBN) and 4- Stacked autoencoder. The performance of each network would be various in the different applications and it is required to be trained carefully based on hyper-parameters tuning such as number of hidden layers, hidden units and learning rate. Fig. 3 shows the first step of our approach:

According to Fig. 3, we represent both labeled (time to event) and unlabeled (censored) instances with together ($X_{train} \bigcup X_{pool}$) to obtain strong representation using pool of unlabeled data. In other words, our framework uses the advantages of abundant unlabeled data to provide less complex and more robust features (labeled and unlabeled) for survival analysis.

In the second step, we apply our novel active learning based survival analysis on the represented/lower dimensions features obtained from the first step. This process is demonstrated in Fig. 4.

According to this Figure, we start by applying a survival analysis method such as Cox-based regression or Random survival forest on represented train set. In the next step we use our novel sampling strategy (explained in the next section) to rank the unlabeled data based on their informativeness level. Then we select the most informative candidate from the pool and add it to the train set and repeat the process until the stop criterion happens. The number of iterations in active learning process is usually based on a stop criterion which could be a fixed number of iterations or a convergence condition. For instance, the iteration process can be stopped when the performance improvement is under a specific threshold.

4.1. Expected Performance Improvement (EPI) Sampling (Query) Strategy

All active learning scenarios as well as pool-based active learning use the informativeness measure for evaluation of unlabeled instances to select the best query (the most informative unlabeled instance). There exist several proposed approaches which formulate such query strategies in the literature which can be categorized in six general frameworks (Settles, 2010): 1- uncertainty sampling, 2- query by committee, 3- expected model change, 4- expected error reduction, 5- variance reduction and 6- density weighted methods.
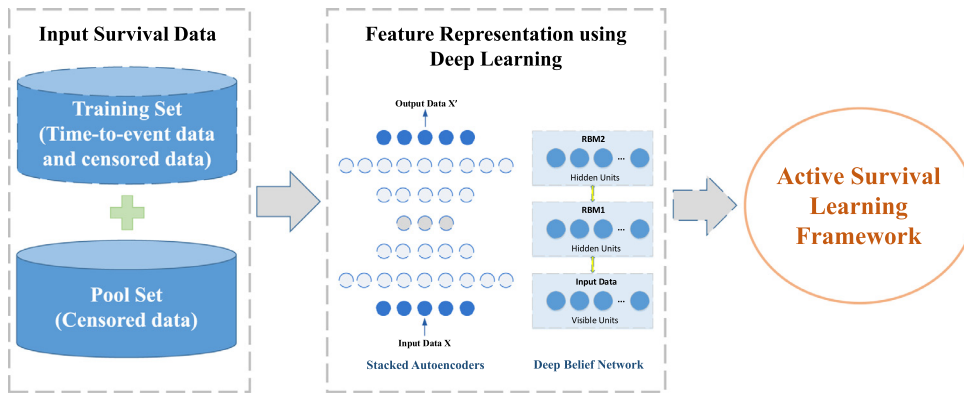
In this research we developed a new sampling (query) strategy based on properties of survival analysis. In our strategy, we select the unlabeled instance as the most informative instance (the

Table 1
Summary of research works used deep learning or active learning in survival analysis.

Authors	Research	DL	AL	SA
Zhou et al. (2013)	Proposed semi-supervised sentiment classification algorithm	✓	✓	
Vinzamuri et al. (2014)	Developed survival regression for censored data for electronic health records		✓	✓
Ranganath et al. (2016)	Introduced a deep hierarchical generative approach for survival analysis in heart disease	✓		✓
Nie, Zhang, Adeli, Liu, and Shen (2016)	Proposed a survival analysis model applied on high-dimensional multi-modal brain images	✓		✓
Liao and Ahn (2016)	Proposed a survival analysis framework using a LSTM model	✓		✓
Huang, Zhang, and Xiao (2017)	Developed a survival model using CNN-based and one FCN-based sub-network and applied on pathological images and molecular profiles	✓		✓
Chaudhary, Poirion, Lu, and Garmire (2017)	Introduced a DL based, survival model on hepatocellular carcinoma patients using genomic data	✓		✓
Liu et al. (2017)	Proposed an active learning approach using DBN for classification of hyperspectral images	✓	✓	
Luck et al. (2017)	Developed a patient-specific kidney graft survival model using principle of multi-task learning	✓		✓
Sener and Savarese (2017)	Developed an active learning framework using CNN for image processing applications	✓	✓	
Katzman et al. (2018)	Proposed a Cox proportional hazards deep neural network for personalized treatment recommendations	✓		✓
Lee et al. (2018)	Developed a survival model using deep learning which trained based on a loss function that uses both risks factors and survival times	✓		✓

Note: DL, AL and SA refer to deep learning, active learning and survival analysis.

Fig. 3. Deep representation of survival data.

best query) when it has the greatest performance change to the current survival model if we knew its label. Our sampling model use concordance index (C-index) to define the informative measure to query the unlabeled data. The survival model is trained again by adding a new instance (X^+) from the pool to the training set: $Train_{new} = Train \bigcup X^+$ and the performance change is formulated based on the c-index difference as follows:

$$\Delta C_{X^+} = C_{new\ model} - C_{current\ model} \tag{6}$$

Similar to the other active learning sampling strategy, our goal is to select the most informative instance which could maximally improve the current model performance. This selection can be formulated as follows:

$$X^* = \underset{X^+ \in pool}{\operatorname{argmax}} \Delta C_{X^+} \tag{7}$$

Since in the real-world applications, we do not know the true label (time to event) of the instances in the pool, we should calculated the expected performance change over all possible time to events (T_s) for each unlabeled records as follows:

$$X^* = \underset{X^+ \in pool}{\operatorname{argmax}} \frac{\sum_{s=1}^{S} h(T_s|X^+)\ \Delta C_{X^+}}{\sum_{s=1}^{S} h(T_s|X^+)} \tag{8}$$

Our sampling strategy works for all survival analysis approaches such as cox-based models, parametric models and random survival forests. As an example for the cox regression, ΔC_{X^+} can be formulated as following equation and X^* is chosen based on Eq. (8).

$$\Delta C_{X^+} = \frac{1}{N} \left[\sum_{\delta_i=1} \sum_{T_i < T_j} (\hat{\beta}_2^s X_i > \hat{\beta}_2^s X_j) - \sum_{\delta_i=1} \sum_{T_i < T_j} (\hat{\beta}_1 X_i > \hat{\beta}_1 X_j) \right] \tag{9}$$

Where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the estimated cox model coefficients trained based on the current and new training set ($Train_{new}$). N refers to the comparable (permissible) pairs in validation set for calculating c-index.

4.2. Proposed Deep Active Survival Analysis (DASA) algorithm

Algorithm 1 describes our Deep Active Survival Analysis approach called DASA in detail. First, we need to set the train and pool data with considering that all instances in the pool should be censored (unknown time-to-event instances). Afterwards, in line 2, we apply deep feature learning on both train and pool sets. This step is an unsupervised learning process to build robust features with lower dimensions from original ones. Deep network should be trained based on hyper-parameters tuning (e.g., learning rate, batch size and hidden units). In this step we need to keep the weights of trained deep network for representation learning of new instances in testing process, it means each instance in the test set should be transformed in the lower dimensions based on these weights. After representation learning, we partition represented in-
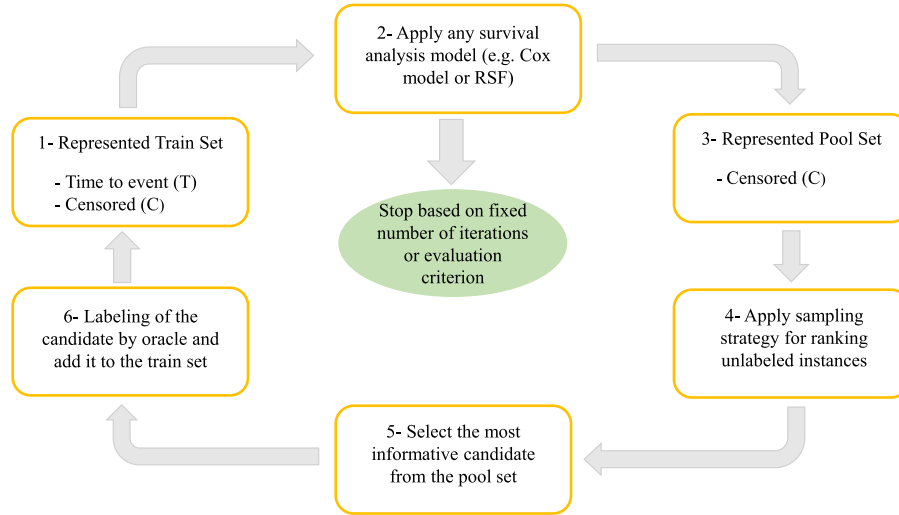
Fig. 4. Active survival analysis approach.

Algorithm 1 Deep Active Survival Analysis (DASA) algorithm.

Require: Training set (X_T), Pool set (X_P), Survival status (δ), Time to event (T), Deep architecture parameters (hidden layers, hidden units, …), Active learning maximum iteration (max_iter)

1: Round = 1
2: Training deep network for feature reduction on ($X_T \bigcup X_P$)
3: $Train\ set \longleftarrow X'_T$
4: $Pool\ set \longleftarrow X'_P$
5: **repeat**
6:     Model = $Deep\_Survival$ (X'_T, δ, T)
7:     **for** each record in the pool ($x \in X'_P$) **do**
8:       Apply EPI sampling strategy and calculate the expected performance improvement for each instance
9:     **end for**
10:     $X^* = argmax_{x \in X'_P} \dfrac{\sum_{s=1}^{S} h(T_s | x)\ \Delta C_x}{\sum_{s=1}^{S} h(T_s | x)}$
11:     Labeling (time-to-event) of X^* by an Oracle based on original features
12:     $X'_P \longleftarrow X'_P - \{X^*\}$
13:     $X'_T \longleftarrow X'_T \bigcup \{X^*\}$
14:     $\delta_{X^*} \longleftarrow 1$
15:     $Round \longleftarrow Round + 1$
16: **until** $Round \neq max\_iter$ (OR reach to convergence condition)

---

stances into train (X'_T) and pool set (X'_P) corresponding to their indexes in the original train and pool sets (lines 3 and 4). These represented sets are considered as the input of active survival model.

In line 6, we start active learning process. First, we apply survival analysis on deep represented features ($Deep\_Survival$). This framework is flexible and all survival models can be used in this step. It is expected the performance of survival model is improved by using represented features (instead of original ones), this improvement is termed as *Deep Learning effect* in this study.

In line 7, we apply our sampling strategy. In each iteration, we calculate the expected performance (ΔC_x) for all instances in the pool and select the best candidate based on our EPI sampling strategy described in the previous section (Eqs. (6)–(8)). Afterward, we label the best candidate using an oracle and remove it from the

pool and insert it to the train set. Then we return to line 6 and apply the survival analysis model on new train set and then repeat the process until reach to maximum number of iteration or a convergence condition (line 16). In each iteration of active learning process, it is expected to improve the performance of survival model which is termed as *Active Learning effect* in this study.

4.3. Treatment recommendations using proposed DASA approach

In this section, we propose a simple yet effective approach to discover treatment patterns and treatment recommendations using DASA. Our method is highly useful when EHRs are high-dimensional and small size. Suppose $X_T = \{X_1^T, X_2^T, \ldots, X_n^T\}$ is the treatment set and $X_A = \{X_1^A, X_2^A, \ldots, X_N^A\}$ refers to all other personalized features related to each patient where $N >> n$. Therefore, the input features is the union of these two sets ($X_T \bigcup X_A$). Since in the case of high-dimensional features, traditional approaches such as Cox proportional hazard or random survival forests cannot find the pattern of specific features (e.g., small treatment set), we first represent X_A using deep learning to a lower dimension set (X'_A) and then combine this represented set with the treatment set (X_T) to build the new feature set ($X_{new} = X'_A \bigcup X_T$). In the second phase, we apply our active learning framework to train an accurate survival model based on new features and then find the pattern of treatment sets and interpret the results (e.g., comparison the coefficient of treatment options using Cox model or finding the importance of different treatment plan using random survival forests).

In our treatment recommendation approach, we transform many clinical features to a small feature set with higher level abstraction and more robust features. While we represent patient information to lower dimension using deep learning we combine non-represented treatment options (as features of interest) with the represented set and then perform survival analysis using active learning framework. In the next section, we demonstrate how our approach discovers the treatment patterns better than traditional approaches.

5. Experimental study: survival analysis for prostate cancer (SEER-Medicare data)

In this section, we evaluate the performance of our approach (DASA) through experimental study. We use the Surveillance, Epidemiology and End Results (SEER)-Medicare linked database from SEER program of the National Cancer Institute (NCI). SEER-
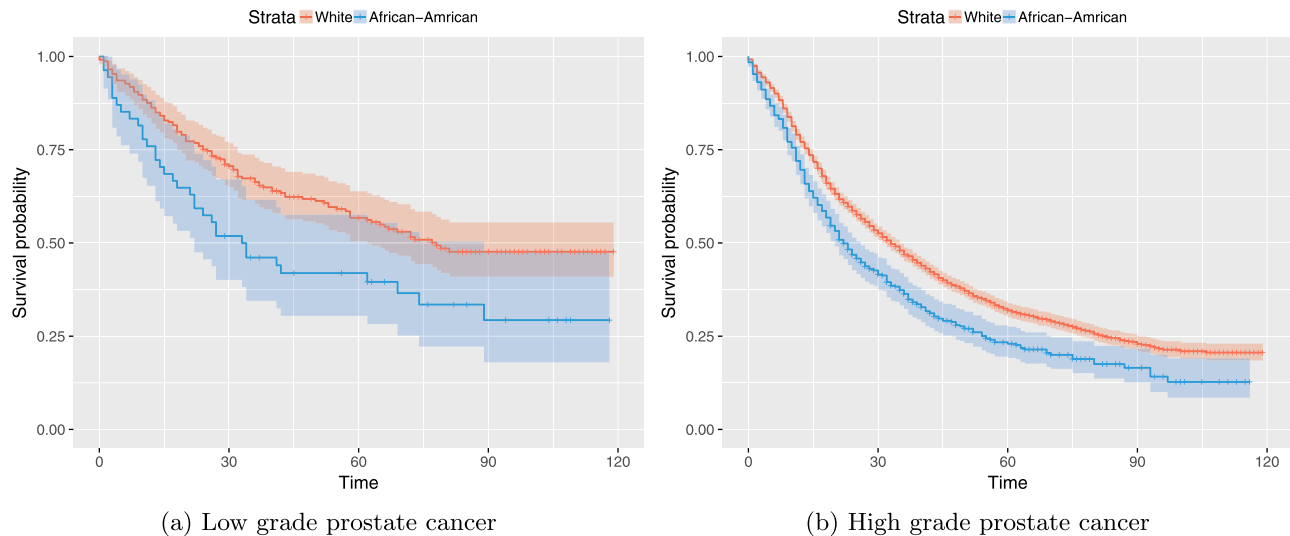
(a) Low grade prostate cancer   (b) High grade prostate cancer

Fig. 5. Kaplan–Meier plots for African–American and white patients.

Medicare data is a powerful and unique source of epidemiological data on the occurrence and survival rates of cancer in the United States. In our study, we use prostate cancer SEER-Medicare data to evaluate our survival analysis approach and provide some insights by treatment recommendation.

5.1. Datasets: SEER-Medicare prostate cancer data

Prostate cancer is the most popular diagnosed invasive cancer among men, with approximately 56% of all prostate cancer patients diagnosed in men with age 65 years and older (Siegel, Miller, & Jemal, 2015). Fortunately, a wide range of men (nearly 90%) are diagnosed with non-metastatic prostate cancer and 5-year relative survival rate is very high for them. The death rate for prostate cancer is different among different populations. A good example of this racial disparity is the death rate for African-American men which is 2.5 times higher than white men. there exists a critical need to develop precision survival analysis for each cohort and discover the pattern of treatment.

5.2. DASA Pperformance for SEER-Medicare prostate cancer data

In this study, we consider the SEER-Medicare data into two racial groups: (1) African–American patients and (2) White patients. Both groups are including many features (more than 300 features) such as demographic data, socioeconomic variables, tumor information and assigned treatment. First, we employed some pre-processing methods to clean the data, for example we removed the features with more than 30% missing values and predict the amount of missing values for otherwise using Random Forest. Also, we applied Isolatin Forest (Liu, Ting, & Zhou, 2008) for anomaly detection and outlier detection. After cleaning process, approximately 1000 and 5000 patients remained for African-American and white patients respectively. For better understanding of this survival data, we demonstrated the Kaplan–Meier curve (Kaplan & Meier, 1958) for low grade (Gleason score = 6) and high grade (Gleason score ≥ 8) prostate cancer in Fig. 5. Kaplan–Meier curve is the most popular plot for survival analysis which indicates the survival probability at the specified survival time. As shown, generally, the survival probability for white patients is higher than African-Americans, and it is decreased in higher grade cancer based on specified survival time for both racial groups.

In our study, the instances are patients diagnosed with prostate cancer and the labels are survival time to event of interest (patient death). While this event has been already happened for some patients, there are many patients without known labels (censored instances in the pool). In our framework we estimate the labels of patients in the pool using an oracle and select the query which is the most informative instance in the pool.

Since SEER-Medicare data is high-dimensional, sparse and complex, feature representation using deep learning can build more robust features when we use pool of unlabeled data (censored instances) in the representation process. In the other hand, our method using active learning has highly potential to improve the performance of survival models when we deal with small sample size (including time-to-event and censored instances). In this way, in experimental study, we consider small set for training of survival model and show the performance of our approach in comparison with baseline.

For labeling of the censored instances (unlabeled data) in active learning framework, we use some scientific reports such as SEER cancer statistics review from National Cancer Institute (NCI) (Howlader et al., 2014) which acts as an oracle (prior knowledge) to estimate the time-to-event (label) of censored instances. One of these statistics is illustrated in Table 2. Since we use prior knowledge to label the censored data, our sampling strategy selects the instances from the pool which could be labeled more accurately. Intuitively, since our approach uses the performance improvement (C-index improvement) in query selection, it selects the instances with more accurate relative risk in comparison to the instances with known labels.

To evaluate the performance of our approach, we first employ CPH regression model (as a well-known survival analysis approach) and demonstrate how DASA can improve its performance based on different training sample size. For deep feature representation we used Stacked Autoencoder (SAE) with 5 hidden layers. Stacked autoencoder is deep architecture constructed by multiple autoencoders. An autoencoder is a shallow network with one hidden layer where the number of unites are the same in the input and output layers. A stacked autoencoder can be trained based on greedy layer-wise approach (Bengio, Lamblin, Popovici, & Larochelle, 2007) which means each autoencoder should be trained by encoding and decoding process one-by-one to minimize the reconstruction error in deep network.

While stacked autoencoder with promised performance is an appropriate choice for deep representation of medical data (Mamoshina et al., 2016), our method is flexible and can use any other deep architectures for survival feature learning.

Table 2
5-Year SEER conditional relative prostate cancer survival and 95% confidence intervals.

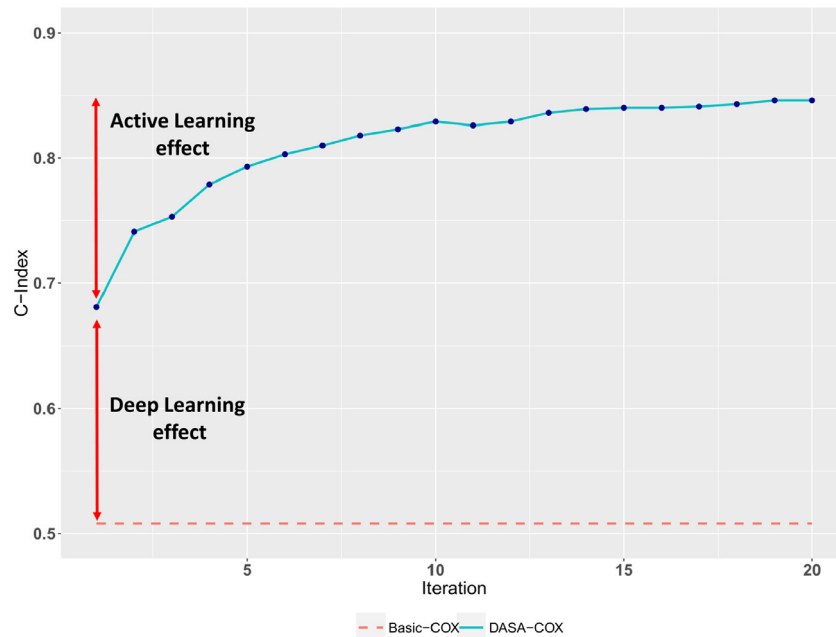| Stage at diagnosis	Survival time since diagnosis	Percent surviving next 5 years	
		Percent	Confidence interval
Local	0-year	100%	(100, 100)
	1-year	100%	(100, 100)
	3-year	100%	(100, 100)
Regional	0-year	100%	(100,100)
	1-year	99.3%	(98.9, 99.5)
	3-year	98.9%	(98.4, 99.2)
Distant	0-year	29.2%	(28.4, 29.9)
	1-year	34.1%	(33.1, 35.1)
	3-year	45.6%	(43.9, 47.2)
Unstaged	0-year	76.6%	(75.6, 77.5)
	1-year	81.1%	(79.8, 82.1)
	3-year	82.8%	(81.4, 84.1)

Fig. 6. Performance of proposed approach in comparison with baseline (training size = 25).

Fig. 6 shows the average performance of our approach for 20 iterations in comparison with baseline on the test data. We randomly selected 50 instances from African-American patients dataset and considered 25 instances as train set and 25 instances as test set, then implemented our approach in 20 iterations. For more robust results, we proceeded this process over 10 runs and calculated the average performance in each iteration. We chose 20 iterations in active learning process since according to our implementations the amount of performance improvement is not remarkable after 20 iterations.

As demonstrated in Fig. 6, our method (DASA-COX) improves the performance of Basic-COX significantly in terms of concordance index. This improvement is caused by two effects: (1) Deep learning effect which improves the model performance by features representation using labeled and unlabeled instances, and (2) Active learning effect which increases the model performance by involving the best labeled censored instance from the pool set in training process across all iterations.

According to our results, the choice of deep architecture plays a significant role in the amount of deep learning effect. In this way, we considered several SAE architectures with different parameters and then applied parameter tuning for major parameters such as learning rate, number of unites in the hidden layers, activation

functions and batch size to select the best parameters in the SAE network. Finally, we selected the best one with 150, 100 and 5 hidden units in the encoders, decoders and latent layers. Since in autoencoders the middle layer provides the highest representation of the input (Bengio et al., 2007), we used the transformed features in the middle layer as the input of survival analysis model.

Fig. 7 shows our approach performance for training size of 50 and 100 instances (we followed the same procedure as previous). Top panel belongs to African–American patients and bottom panel is related to white patients. It is clear DASA-COX outperforms baseline approach in all cases. The effect of deep learning in improving model performance is higher at the bottom panel which is caused by larger amount of pool set related to white patients that provides better feature learning. As mentioned before, our approach is flexible enough and can employ any survival analysis model in its framework to improve the baseline. Hence, we perform Random Survival Forests (RFS) model as a well-known non-linear survival model along with CPH model and evaluate our approach across different training sizes. The results are shown in Tables 3 and 4 for African-Americans and white patients respectively.

The results confirm that our method can improve the concordance index significantly for Cox proportional hazard model and random survival forests across all training sizes in each dataset.
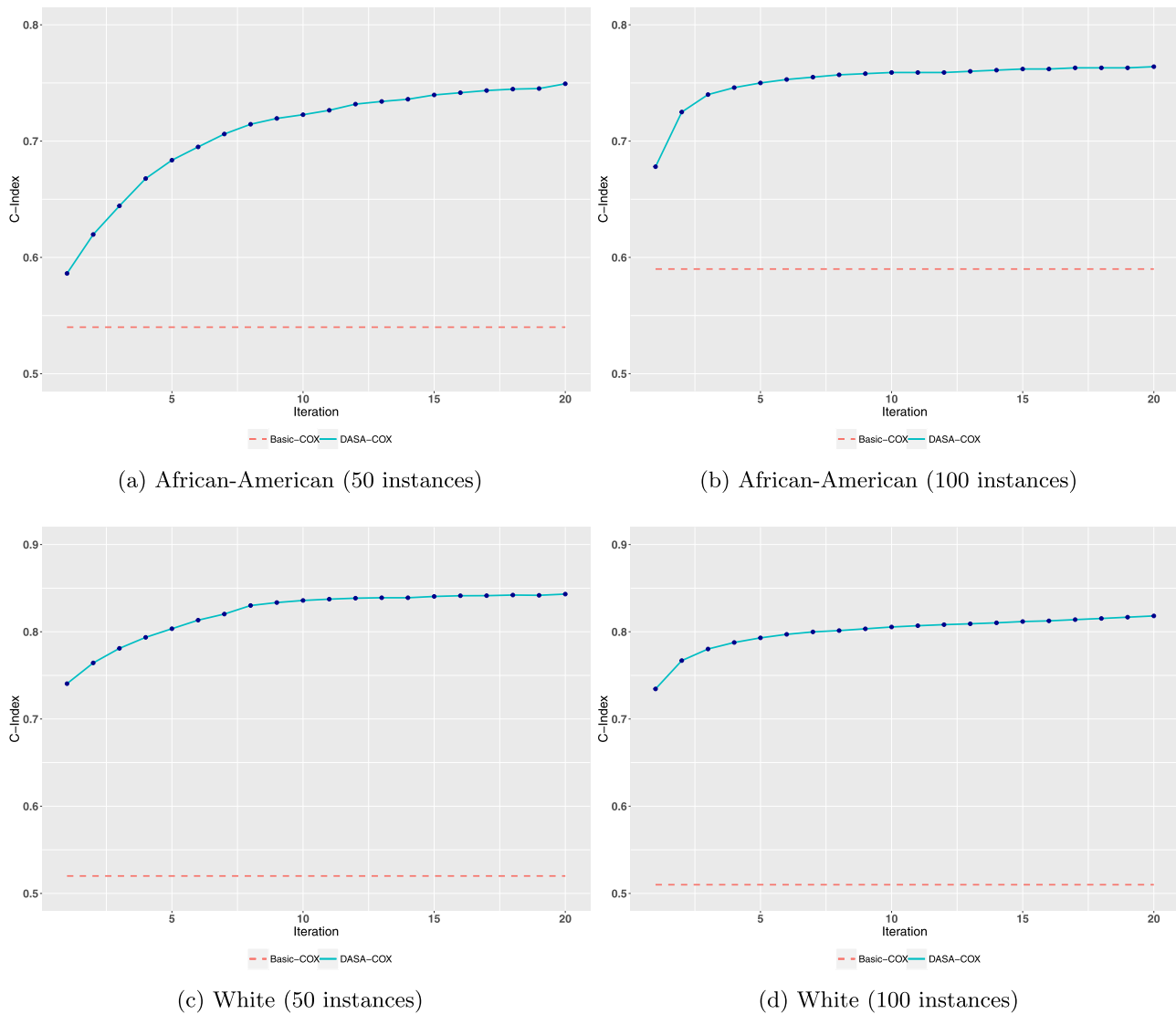
(a) African-American (50 instances)

(b) African-American (100 instances)

(c) White (50 instances)

(d) White (100 instances)

Fig. 7. Performance of proposed approach in comparison with baseline for different training size.

Table 3
Performance comparison (C-index) between DASA and baseline models (African-Americans).

Training size	CPH	DASA-CPH	RSF	DASA-RSF
25 instances	55.2%	**84.7%**	16.3%	**57.6%**
50 instances	54.2%	74.9%	17.6%	54.5%
100 instances	59.1%	76.6%	21.4%	48.2%
200 instances	58.6%	72.6%	22.3%	47.9%

Table 4
Performance comparison (C-index) between DASA and baseline models (Whites).

Training Size	CPH	DASA-CPH	RSF	DASA-RSF
25 instances	52.4%	**87.9%**	13.3%	**62.1%**
50 instances	51.2%	84.4%	15.5%	58.3%
100 instances	50.8%	82.3%	15.7%	49.7%
200 instances	53.6%	77.1%	18.2%	46.4%

According to above results, the largest performance improvement for both methods (CPH and RSF) is achieved in the 25 instances training size (the smallest one) caused by active learning effect. In fact, we can conclude that DASA leads to larger performance im-

provement in smaller training size where the active learning effect is dominant. Active learning achieves larger improvement in the smaller dataset (Settles, 2010) where the performance of baseline model is lower, therefore it can make larger impact (difference from baseline) in few iterations.

5.3. Evaluation of EPI sampling strategy

To evaluate our novel sampling (Expected Performance Improvement) strategy, we consider two popular sampling strategies for active learning (Settles, 2010): (1) Random sampling, where the queries are randomly selected from the pool and (2) Uncertainty sampling strategy, in this case an active learner selects the queries of instances which are least certain how to label. We applied each sampling strategies (EPI, Random and Uncertainty) in DASA framework using CPH. The results indicate that our EPI sampling strategy outperforms the other ones in both datasets (Tables 5 and 6).

5.4. Treatment recommendation by DASA

In the second step, we demonstrate how our treatment recommendation approach works. we considered three well-known treatment options for prostate cancer: chemotherapy, radiotherapy

Table 5
Performance comparison of different sampling strategies (African–Americans).

Training size	EPI	Random	Uncertainty
25 instances	**84.7%**	65.4%	77.1%
50 instances	74.9%	62.6%	70.4%
100 instances	76.6%	58.6%	71.9%
200 instances	72.6%	61.4%	69.3%

Table 6
Performance comparison of different sampling strategies (Whites).

Training size	EPI	Random	Uncertainty
25 instances	**87.9%**	67.5%	82.8%
50 instances	84.4%	63.3%	78.3%
100 instances	82.3%	64.2%	77.6%
200 instances	77.1%	60.8%	65.6%

Table 7
Average hazard ratio among different treatment plans.

	Method	Chemotherapy	Radiotherapy	Surgery
African–Americans	COX-Base	1	1	1
	COX-DASA	0.74	1.04	1.38
White patients	COX-Base	1	1	1
	COX-DASA	0.96	1.08	2.23

Table 8
Average hazard ratio among different treatment plans.

	Cancer type	Chemotherapy	Radiotherapy	Surgery
African–Americans	Low grade	0.75	0.59	0.88
	High grade	0.53	0.96	1.27
White patients	Low grade	1.86	0.65	1.13
	High grade	0.55	1.64	2.63

high cancer grade (Gleason score is higher than 7), the results of risk associated with different treatment options (obtained from COX-DASA) are shown as following:

According to the Table 8, the risk (hazard ratio) of treatment options are different based on the grade of cancer in each racial group. For example treatment with surgery in the high grade prostate cancer is more risky rather than low grade cancer which is confirmed in the literature as well (Carter, 2011; Erickson et al., 2018; Lei et al., 2015). Based on this results, among all groups, the risk of surgery for white patients with high grade cancer is very large in comparison with the other two treatment options.

This experimental treatment recommendation was a simple example to show how our method works. This approach is highly useful for comparing the risk associated with new treatment in comparison with current treatment plans where the labeled data is rare and expensive.

6. Discussion and conclusion

In this research, we proposed a novel survival analysis framework using deep learning and active learning called Deep Active Survival Analysis (DASA). The motivation for this study comes from either literature gap and application needs in several domains (e.g., healthcare, manufacturing and finance) where the labeled data is scarce and high-dimensional.

Our approach is able to improve the survival analysis performance significantly and provides treatment recommendations. DASA encompasses two main phases: (1) deep feature learning and (2) active learning process. We do feature representation using deep learning to produce robust features from high-dimensional, sparse and complex EHRs. We used the advantage of pool of unlabeled data (censored instances) to provide better representation of labeled instances from deep learning implementation. In the active learning process, we developed a new sampling strategy specifically for survival analysis which can be used for any survival analysis models such as Cox-based approaches and random survival forests.

In the experimental study, we used SEER-Medicare data related to prostate cancer among African-Americans and white patients to demonstrate how our model can enhance the performance of survival analysis in comparison of traditional approach. Empirically we showed that deep learning has greater effect on survival performance improvement in the case that we have larger pool of unlabeled data (because of stronger unsupervised feature learning) and active learning effect is higher when we deal with smaller training sample size (because learning from smaller data is hard and active learning can help more). We applied our treatment recommendation approach to find hazard ratio of three common treatment plan (chemotherapy, radiotherapy and surgery) for prostate cancer based on Cox model. While traditional CPH model fails to find the hazard ratios among high dimensional data, our approach discovers them and provides some racial treatment insights, for example, surgery is associate with higher risk in white patients especially in the case of high grade prostate cancer. We also evaluated our new sampling strategy (EPI) by comparing with two popular sampling strategies in active learning literature. The results showed that our strategy outperforms the others.

and surgery as three binary variables in our dataset. Our goal is to discover the importance of each therapy using DASA approach for each subgroup of patients (African–Americans and white patients). Since in the experimental study CPH illustrated a great performance, we performed survival analysis using CPH. We do feature representation by deep stacked autoencoder network with 150, 100 and 5 hidden unites in encoder, decoder and latent layers respectively. Without loss of generality, we used small sample size with 50 instances in the training process (this number of instances could be various and this is just an example to show how our recommendation approach works). Before training process, we combined chemotherapy, radiotherapy and surgery variables (features of interest) to the represented features in both training and pool datasets. The represented features are the results of deep feature learning performed on other features (high-dimensional features in training and pool sets). In our case, we combined the 5 represented features obtained from the latent layer of SAE network with the three features of interest. Afterward, we trained the cox survival model using active learning framework with 20 iterations over all new features (totally 8 features). The results for average of exponential of coefficients (hazard ratios) over 10 runs shown in Table 7 for African-Americans and white patients:

As shown above, traditional CPH model could not differentiate between treatment plans where their hazard ratios are one. Since the data is high-dimensional traditional CPH leads to zero coefficients for these three treatment variables. On the other side, our approach using Cox model can discover the risk associated to each treatment. Based on our results, surgery has the highest risk in the both subgroup of patients, radiotherapy is associate with a decline in the survival rate while chemotherapy increases the survival rate with lowest risk. It is obvious that the pattern of hazard ratios among treatment plans are different between African-American and white patients. For example the risk related to surgery is significantly higher than the other two therapies in white patients (more than 2 times) while in the African-Americans the pattern is different.

For more evaluation of our treatment recommendation approach, we considered two groups of patients: (1) patients with low cancer grade (Gleason score equal to 6) and (2) patients with

In sum, our method leads to more accurate survival analysis for risk prediction, survival time estimation and treatment recommendation. Our approach is flexible enough to capture any survival analysis model and improves its performance. Our model can be applied on different areas especially in the case of testing and comparing the risk (impact) of new treatment (e.g., in healthcare) or new technology (e.g., in the manufacturing process). In this way, the most important challenge is providing a reliable oracle in active learning process because in this case the prior knowledge is limited and taking the advantages of experts is expensive and time consuming.

There are some limitations to the proposed framework, the quality of the results are limited by the cases where data is high dimensional and labeled instances are small. Although we used the rich SEER-Medicare prostate cancer data to evaluate our approach, more experimental studies are needed to acknowledge the performance of the proposed approach. For evaluation of our treatment recommendation method, we only considered three recommendation options while it could be extended for the multiple choices of therapies. Although DASA achieved good results, the current performance might be improved by applying more deep architectures for feature learning, developing better sampling strategies and providing more reliable oracle in active learning process.

For the future works, DASA can be applied on the other datasets and applications including real-world cases in the healthcare, manufacturing and finance to provide more evaluations and insights. The sampling strategy proposed in this study is based on the performance improvement which can be enhanced in the other directions. As the choice of deep representation plays a significant role in the success of machine learning tasks (Bengio et al., 2007), it is a great opportunity to investigate and compare the performance of the other deep architectures such as Deep Belief Network and Variational Autoencoders in DASA framework. Last but not least, since the proposed approach uses deep learning and active learning as an integrated framework, it is essential to evaluate its computational performance when we deal with big unlabeled data in the pool and develop appropriate strategies if needed.

Acknowledgment

References

Angluin, D. (1988). Queries and concept learning. *Machine Learning, 2*(4), 319–342.

Atlas, L. E., Cohn, D. A., & Ladner, R. E. (1990). Training connectionist networks with queries and selective sampling. In *Advances in neural information processing systems* (pp. 566–573).

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(8), 1798–1828.

Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2007). Greedy layer-wise training of deep networks. In *Advances in neural information processing systems* (pp. 153–160).

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends® in Machine Learning, 2*(1), 1–127.

Carter, H. B. (2011). Management of low (favourable)-risk prostate cancer. *BJU International, 108*(11), 1684–1695.

Chaudhary, K., Poirion, O. B., Lu, L., & Garmire, L. X. (2017). Deep learning based multi-omics integration robustly predicts survival in liver cancer. *Clinical Cancer Research*, clincanres-0853.

Cox, D. R. (1992). Regression models and life-tables. In *Breakthroughs in statistics* (pp. 527–541). Springer.

Erickson, A., Sandeman, K., Lahdensuo, K., Nordling, S., Kallajoki, M., Seikkula, H., et al. (2018). New prostate cancer grade grouping system predicts survival after radical prostatectomy. *Human Pathology, 75*, 159–166.

Heagerty, P. J., & Zheng, Y. (2005). Survival model predictive accuracy and ROC curves. *Biometrics, 61*(1), 92–105.

Hosmer, D. W., Lemeshow, S., & May, S. (2011). *Applied survival analysis*. Wiley Blackwell.

Howlader, N., Noone, A., Krapcho, M., Garshell, J., Neyman, N., Altekruse, S., et al. (2014). *Seer cancer statistics review (CSR) 1975–2011*. Bethesda, MD: National Cancer Institute. 2014

Hsu, D. Algorithms for Active Learning. PhD thesis, Department of Computer Science and Engineering, School of Engineering, University of California, San Diego, 2010.

Huang, C., Zhang, A., & Xiao, G. (2017). Deep integrative analysis for survival prediction.

Ishwaran, H., Gerds, T. A., Kogalur, U. B., Moore, R. D., Gange, S. J., & Lau, B. M. (2014). Random survival forests for competing risks. *Biostatistics, 15*(4), 757–773.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association, 53*(282), 457–481.

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). Deep-Surv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Medical Research Methodology, 18*(1), 24.

Kleinbaum, D. G., & Klein, M. (2010). *Survival analysis*: 3. Springer.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.

Lee, C., Zame, W. R., Yoon, J., & van der Schaar, M. (2018). DeepHit: a deep learning approach to survival analysis with competing risks.

Lei, J. H., Liu, L. R., Wei, Q., Yan, S. B., Song, T. R., Lin, F. S., et al. (2015). Systematic review and meta-analysis of the survival outcomes of first-line treatment options in high-risk prostate cancer. *Scientific Reports, 5*, 7713.

Lewis, D. D., & Gale, W. A. (1994). A sequential algorithm for training text classifiers. In *Proceedings of the 17th annual international ACMSIGIR conference on research and development in information retrieval* (pp. 3–12). Springer-Verlag New York, Inc..

Liao, L., & Ahn, H.-i. (2016). Combining deep learning and survival analysis for asset health management. *International Journal of Prognostics and Health Management*.

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation forest. In *2008 Eighth IEEE international conference on data mining* (pp. 413–422). IEEE.

Liu, P., Zhang, H., & Eom, K. B. (2017). Active deep learning for classification of hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10*(2), 712–724.

Luck, M., Sylvain, T., Cardinal, H., Lodi, A., & Bengio, Y. (2017). Deep learning for patient-specific kidney graft survival analysis. arXiv preprintarXiv:1705.10245.

Mamoshina, P., Vieira, A., Putin, E., & Zhavoronkov, A. (2016). Applications of deep learning in biomedicine. *Molecular Pharmaceutics, 13*(5), 1445–1454.

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2017). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics*, bbx044.

Nezhad, M. Z., Zhu, D., Sadati, N., & Yang, K. (2018). A predictive approach using deep feature learning for electronic medical records: A comparative study. arXiv preprintarXiv:1801.02961.

Nie, D., Zhang, H., Adeli, E., Liu, L., & Shen, D. (2016). 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. In *International conference on medical image computing and computer-assisted intervention* (pp. 212–220). Springer.

Ranganath, R., Perotte, A., Elhadad, N., & Blei, D. (2016). Deep survival analysis. arXiv preprintarXiv:1608.02158.

Ravi, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., et al. (2017). Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics, 21*(1), 4–21.

Sener, O., & Savarese, S. (2017). A geometric approach to active learning for convolutional neural networks. arXiv preprintarXiv:1708.00489.

Settles, B. (2010). *Active learning literature survey* p. 11. University of Wisconsin, Madison. 52

Shickel, B., Tighe, P., Bihorac, A., & Rashidi, P. (2017). Deep EHR: A survey of recent advances on deep learning techniques for electronic health record (EHR) analysis. arXiv preprintarXiv:1706.03446.

Siegel, R. L., Miller, K. D., & Jemal, A. (2015). Cancer statistics, 2015. *CA: ACancer Journal for Clinicians, 65*(1), 5–29.

Vinzamuri, B., Li, Y., & Reddy, C. K. (2014). Active learning based survival regression for censored data. In *Proceedings of the 23rd ACM international conference on conference on information and knowledge management* (pp. 241–250). ACM.

Wang, P., Li, Y., & Reddy, C. K. (2017). Machine learning for survival analysis: A survey. arXiv preprintarXiv:1708.04649.

Zhou, S., Chen, Q., & Wang, X. (2013). Active deep learning method for semi-supervised sentiment classification. *Neurocomputing, 120*, 536–546.