Available online at www.medicinescience.org

ORIGINAL ARTICLE

Medicine Science International Medical Journal

Medicine Science 2022;11(3):1202-6

Comparison of Performance of Deep Survival and Cox Proportional Hazard Models: an Application on the Lung Cancer Dataset

Kubra Elif Akbas[1], Ipek Balikci Cicek[2], Mehmet Onur Kaya[1], Cemil Colak[2]

[1]Fırat University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Elazıg, Turkey
[2]Inonu University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Malatya, Turkey

Received 31 March 2022; Accepted 22 June 2022
Available online 08.08.2022 with doi: 10.5455/medscience.2022.03.078

Abstract

The goal of this study is to compare the performance of the deep survival model and the Cox regression model in an open-access Lung cancer dataset consisting of survivors and dead patients. In the study, it is applied to an open access dataset named "Lung Cancer Data" to compare the performances of the CPH and deepsurv models. The performance of the models is evaluated by C-index, AUC, and Brier score. The concordance index of the deep survival model is 0.64296, the Brier score was 0.128921, and the AUC was 0.6835. With the Cox regression model, the concordance index is calculated as 0.61445, brier score 0.1667, and AUC 0.5832. According to the Concordance index, brier score, and AUC criteria, the deep survival model performed better than the cox regression model. DeepSurv's forecasting, modeling, and predictive capabilities pave the path for future deep neural network and survival analysis research. DeepSurv has the potential to supplement traditional survival analysis methods and become the standard method for medical doctors to examine and offer individualized treatment alternatives with more research.

Keywords: Cox regression, deep survival, survival, deep learning

Introduction

Survival analysis is the analysis of time data from the beginning of an event of interest to the endpoint [1]. Survival analysis is a fundamental research tool utilized in a wide range of domains, including medicine, biology, epidemiology, demographics, and engineering [2]. Survival analysis is vital in medical prognosis and risk assessment [3]. Medical researchers use survival models to determine the significance of predictive variables in outcomes like death and cancer recurrence, and then inform patients about treatment alternatives. The Cox proportional hazards model (CPH) is a semi-parametric model that calculates the effect of observed covariates on the likelihood of an event occurring [4,5]. The death risk of a patient is assumed to be a linear function of their variables in CPH. This hypothesis is known as the condition of the linearly proportionate hazard. In many real-world datasets, the assumption that the risk function is linear is likely to be oversimplified. A larger family of survival models is required to better fit survival data with nonlinear risk functions [4].

Various algorithms for survival analysis have been created in recent years by academics in the domains of machine learning and data mining. These algorithms can be broken down into two categories: applying machine learning techniques to augment traditional statistical methods and adapting regular machine learning models to handle censored data [6]. To more precisely represent survival data, a survival analysis can combine the benefits of deep neural networks. Katzman et al. introduced the DeepSurv model, which blends deep learning and a multi-layered neural network to create a system for individualized treatment suggestions [7]. Deepsurv is a CPH extension that replaces linear regression in the Cox model with a deep neural network, effectively tackling nonlinearity [6]. With this model, deep learning is used to model the risk function in the hazard function [8].

In this study, deep survival and Cox regression models were applied to an open access Lung cancer dataset consisting of living and dead patients, and it was aimed to compare the performances of the models.

***Corresponding Author:** Kubra Elif Akbas, Fırat University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Elazig, Turkey
E-mail: keet@firat.edu.tr

Materials and Methods

Dataset

In the study, it was applied to an open access dataset named "Lung Cancer Data" [9] to compare the performances of the CPH and deepsurv models. The recommended number of observations per independent variable, to generalize the results, should be 5, but this number is anticipated to be more than 10, particularly between 15 and 20. If the sample accurately represents the population, it is argued that when these numbers are attained, the results can be generalized. The assumption of a multiple of 10 was used to establish the minimum sample size, which was determined to be 100 because there are 10 independent variables [10]. There are 228 patients in the data set used. 165 (72,4%) of these patients died after a certain period of follow-up. The median age of the patients in this study was 63 (39-82).

Cox Proportional Hazard Regression

Survival analysis (follow-up analysis) in medicine, risk of various diseases, prognostic factors (explanatory variables, independent variables, covariates, determinants), treatment success, etc. is of great importance in the examination of data and is applied in many fields of medicine. One of the methods frequently used in survival analysis is the Cox Regression model, which is used to determine the factors that affect life expectancy. The model (1972) is named after its developer, David R. Cox [11].

The Cox regression method is a statistical technique for determining the link between a patient's life expectancy (dependent variable) and several independent variables. The purpose of Cox regression analysis is to build a model that accurately reflects the state of the survival data in general. The impacts of independent variables that are assumed to be effective on life expectancy can be explained and assessed simultaneously in this way. When this model is applied to the lives of patients in a clinical trial, it aids in the separation of treatment effects from the effects of other variables. CPH is more effective than parametric analysis in cases where the assumptions required by parametric models (normality, independence, etc.) are not provided. The following are the basic assumptions of the CPH: (1) The independent variables have a log-linear effect on the hazard function. (2) The risk function has a multiplicative connection with the log-linear function of the independent variables. Aside from these two assumptions, the observations should be independent of one another, and the risk ratio should remain constant throughout time. The proportional risk assumption refers to this assumption regarding the risk ratio [12].

CPH is one of the most used methods to estimate the individual survival curve by taking the hazard function [13]. The survival function and the hazard function are the two main functions of survival analysis. The likelihood of an individual surviving after time t is defined by the survival function, which is denoted by $S(t)=Pr(t>t)$. The following is the definition of the hazard function $\lambda(t)$.

$$\lambda(t) = \lim_{\delta \to 0} \frac{P_r\,(t \le T < t+\delta | T \ge t)}{\delta}$$

The hazard function is the probability that an individual will not survive for an extra infinitesimal time δ, provided that he or she has survived to time t. Therefore, a greater hazard means a greater risk of death. The model presumes that the hazard function consists of two non-negative functions: a principal hazard function $\lambda_0(t)$ and a risk score $r(x)=e^{h}(^{x})$ (It is defined as the effect of an individual's observed covariates on baseline hazard). The hazard function is as follows.

$$\lambda(t \mid x)=\lambda_0\,(t).e^{h(x)}$$

CPH is a proportional hazards function that combines a linear function $\hat{h}_\beta(x)=\beta^T x$ [or equivalently $\hat{r}_\beta(x)=e^{\beta T}x$)] to estimate the log-risk function h_x. The β weights are modified to optimize the Cox partial probability when doing the Cox regression. Given the set of persons still at risk at the time T_i, the partial probability is the product of the likelihood that the event will occur in individual I at each event time T_i. β is the parameter for the Cox partial probability, which is defined as:

$$L_c(\beta) = \prod_{i:E_1=1} \frac{\hat{r}_\beta(x_i)}{\sum_{j \epsilon \Re(T_i)} \hat{r}_\beta(x_j)} = \prod_{i:E_1=1} \frac{exp(\hat{h}_\beta(x_i))}{\sum_{j \epsilon \Re(T_i)} exp(\hat{h}_\beta(x_j))}$$

Here, the T_i, E_i and x_i values are the corresponding event time, event indicator, and the basic data for the ith observation, respectively. The output is defined through the patient group with an observable event $E_1=1$. The risk set $R(t)=\{i:T_i \ge t\}$ is the group of patients still at risk of failure at time t [4].

Deepsurv Network

The first adaptation of survival analysis to neural networks (Farragi and Simon, 1995) was based on a single hidden layer version of the CPH. The initial model's major goal was to discover correlations between primary factors and the hazard risk function. Following the creation of the neural network design using Cox regression, it was discovered that maintaining the main proportionality assumption of the CPH model in real-world huge datasets with non-linear interactions between variables is quite difficult. Farragi and Simon's network, on the other hand, increased the non-linearity quality [14]. J.L. Katzman et al. presented DeepSurv, a more advanced deep learning architecture, as a supplement to Simon-network Farragi's a few years ago. It revealed that when dealing with non-linear data, the CPH model and performance indicators improved [4].

DeepSurv is a deep-feed forward neural network with a large number of layers. It can be used to calculate the influence of variables on survival in patients. This network's structure consists of a huge number of simulated neurons organized into three primary layers: input, hidden, and output. While multiple hidden levels are permitted, just one input and output layer is permitted [7]. DeepSurv is a multilayer network like the Faraggi-Simon network. This network is considered a nonlinear extension of the CPH. However, it allows for a deep architecture (i.e. multiple hidden layers) [5]. It's a deep feedforward neural network that calculates the effects of a patient's covariates on the hazard ratios using the θ network's weights. The main components of deepsurv are illustrated in figure 1. A patient's baseline X data

is their network access. A fully connected node layer, followed by a dropout layer, make up the the network's hidden layers. The $h_0(x)$ network produces a single node with linear activation that estimates the Cox model's log-risk function.

$$l(\theta) := -\frac{1}{N_{E=1}} \sum_{i:E_i=1} \left(\hat{h}_\theta(x_i) - log \sum_{j \epsilon \Re(T_i)} e^{\hat{h}_\theta(x_j)} \right) + \lambda . \|\theta\|_2^2$$

where $N_{E=1}$) is the number of patients with an observable event and λ is the regularization parameter l_2 [4]. The flow diagram of deep survival is given in figure 1.

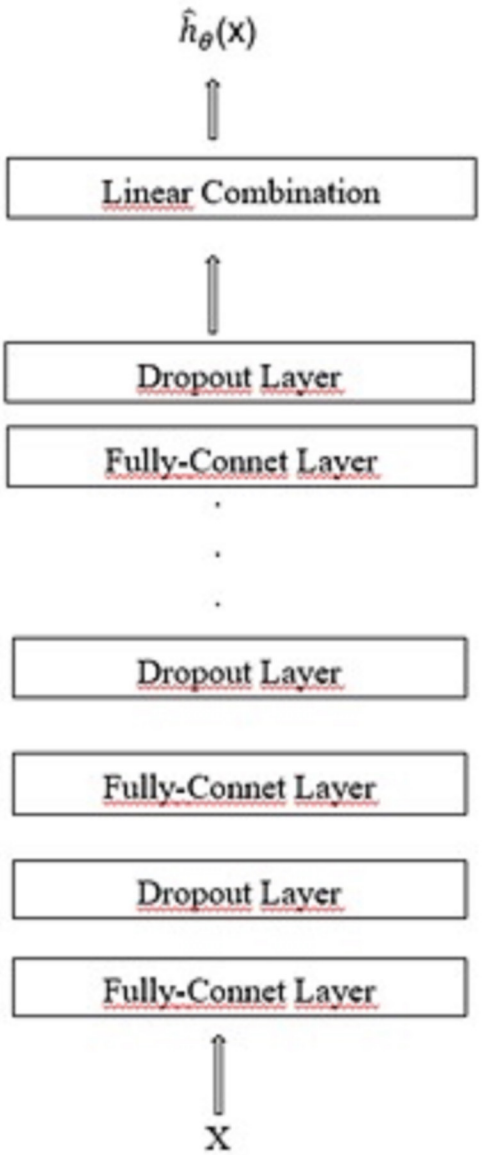

Figure 1. DeepSurv Diagram

Performance Metrics

The concordance index, or C-index, is one of the most extensively used assessment metrics in survival analysis [4,15]. This is likely a result of its interpretability, as classification accuracy and ROC have a close relationship with AUC. The C-index predicts the probability that, for a random pair of individuals, the estimated survival times of two individuals will rank the same as the actual survival times. Because the C-index depends only on the ranking of the estimates, it is very useful for evaluating proportional hazard models. This is because the ordering of proportional hazard models does not change over time. This allows us to use the relative risk function instead of a metric for estimated survival time [15]. The C-index is also a metric for how effectively a model predicts the rank of a patient's death time [4]. This measure evaluates the order of predicted survival times: 0 indicates inconsistency and 1 indicates a perfect fit. [5,16,17]. The Brier score, which measures the average squared distances between the observed survival status and the expected survival probability at a certain time point, with a smaller value indicating better performance, assesses the accuracy of a predicted survival probability at a specific time point [18]. Brier score ranges from 0 to 1, with a lower value indicating better model accuracy [19].

Data Analysis

Quantitative variables were summarized using the median and minimum-maximum methods, whereas qualitative variables were reported using numbers and percentages. Using the Pearson chi-square test, the qualitative variables were compared. For quantitative variables, the Kolmogorov-Smirnov test was employed to verify normality. Comparing quantitative variables between two groups was done using the Mann-Whitney U test. For all tests, the statistical level of significance was set at 0.05. The IBM SPSS Statistics Version 26.0 statistical software package was used for all analyses. Cox regression and deep survival models are analyzed using the Python 3.9.7 programming language [20].

Results

Descriptive statistics for the independent variables in this study are given in Table 2. According to this table, there is a statistically significant difference in time, ph. carno, and pat. carno variables between the groups of the dependent variable (Death events). In terms of inst, age, meal cal., and weight loss variable, however, there was no statistically significant difference between the groups of the dependent variable (Death events) (p>0.05).

The dataset used in the study included a total of 228 patients, 63 living and 165 deceased. The median survival time of the patients was 310±21.77 days. In addition, 1- and 3-year survival were 39% and 6.7%, respectively 1.

Table 1. Distribution table of qualitative independent variables

Variables		Death events		p value*
		Survived(n=63)	Dead(n=165)	
Sex	Male	26(41.3%)	112(67.9%)	<0.001
	Female	37(58.7%)	53(32.1%)	
Ph.ecog	0	26(41.3%)	37(22.6%)	0.006
	1	31(49.2%)	82(50.0%)	
	2	6(9.5%)	44(26.8%)	
	3	0(0%)	1(0.6%)	

*:Pearson chi-square test

Table 1 reveals that the dependent variable (Death events) groups have a statistically significant association with sex and ph. ecog variables (p<0.05).

Table 3 shows that; in survival analysis, the model with a high Concordance index and AUC and a low brier score is better. The concordance index of the deep survival model was 0.64296, the Brier score was 0.128921, and the AUC was 0.6835. With the Cox regression model, the concordance index was calculated as 0.61445, brier score 0.1667, and AUC 0.5832. According to the Concordance index, brier score, and AUC criteria, the deep survival model performed better than the cox regression model.

Table 2. Descriptive statistics for quantitative independent variables

Variables	Death events		p value*
	Survived(n=63) Median(min-max)	Dead(n=165) Median(min-max)	
Inst	12(1-33)	11(1-33)	0.068
Time	284(92-1022)	226(5-883)	**0.003**
Age	62(39-77)	64(40-82)	0.053
Ph. carno	90(50-100)	80(50-100)	**0.005**
Pat. carno	90(30-100)	80(30-100)	**0.004**
Meal cal.	975(96-2450)	1025(169-2600)	0.427
Wt.loss	4(-10-49)	8(-24-68)	0.317

*: Mann Whitney U test

Table 3. Predictive performance (C-index) of the Cox proportional hazard model and DeepSurv

Methods	C-Index	Brier score	AUC
Deep survival	0.64296	0.128921	0.6835
Cox regression	0.61445	0.1667	0.5832

Discussion

Classical analysis methods are insufficient in medical research. The first reason for this is that most of the time, an evaluation of the trial must be done before all patients have died or the outcome under investigation becomes available. Otherwise, it may take years to determine which treatment method is better and the factors affecting the disease. The second reason is that the treatments applied to the patients do not start at the same time. In this type of study, survival analysis methods give more appropriate results [21].

CPH has the feature of being the most used model in survival analysis. In case the effects of the explanatory variables included in the model are not linear, the use of the fractional polynomial Cox regression model is suggested. With the deep survival model, personalized treatment recommendations for patients can be put forward, as well as treatments to be used by doctors to improve patient lives.

Traditional CPH necessitates large cohorts to train models on high-dimensional datasets with many features. As a result, a small number of characteristics are chosen through a subjective process that is prejudice-prone and limited by a lack of information about disease biology. In the machine-learning business, high-dimensional learning issues are widespread, and many machine-learning algorithms have been tweaked to predict survival or progression time. Because of advances in technique, processing hardware, and datasets, neural networks, also known as deep learning, have shattered performance benchmarks in traditional machine learning applications. These networks are made up of densely interconnected layers that use adaptive learning of connectivity parameters to gradually transform inputs into more predictive features [22].

Katzman et al. recently suggested DeepSurv, a deep fully connected network that can be thought of as a nonlinear extension of the CPH. DeepSurv avoids the issues that plague the traditional CPH and outperforms it [4].

This study used an open-access Lung Cancer dataset to compare the performance of the CPH and the deep survival model. According to the findings, the deep survival model's concordance index was 0.64296, the Brier score was 0.128921, and the AUC was 0.6835. The concordance index was computed as 0.61445, the Brier score was 0.1667, and the AUC was 0.5832 using the Cox regression model. The deep survival model outperformed the cox regression model based on concordance index, brier score, and AUC criteria.

Conclusion

As a result, thanks to the model flexibility, using deep learning in survival analysis gives better results. Deepsurv is often as good as or better than the CPH method at predicting patients' risk. DeepSurv's prediction, modeling, and recommendation capabilities open the way for future deep neural network and survival analysis research. With the deep survival model, personalized treatment recommendations for patients can be put forward, as well as the treatments that will be used by doctors to improve patient lives.

Conflict of interests
The authors declare that there is no conflict of interest in the study.

Financial Disclosure
The authors declare that they have received no financial support for the study.

Ethical approval
There is no need for an informed consent form as the open source dataset is used in the study.

References

1. Kartsonaki C. Survival analysis. Diagnostic Histopathology. 2016;22:263-70.
2. Knox KL, Bajorska A, Feng C, et al. Survival analysis for observational and clustered data: an application for assessing individual and environmental risk factors for suicide. Shanghai Arch Psychiatry. 2013;25:183.
3. Thammasorn P, Schaub SK, Hippe DS, et al. Regularizing the deepsurv network using projection loss for med risk assessment. IEEE Access. 2022;10:8005-20.
4. Katzman JL, Shaham U, Cloninger A, et al. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med Res Methodol. 2018;18:1-12.
5. Katzman JL, Shaham U, Cloninger A, et al. Deep survival: A deep cox proportional hazards network. Stat. 2016;1050:1-10.
6. Wang L, Li Y, Chignell M. Combining ranking and point-wise losses for training deep survival analysis models. IEEE International Conference on Data Mining (ICDM). 2021;689-98.
7. Yu H, Huang T, Feng B, Lyu J. Deep-learning Model for predicting the survival of rectal adenocarcinoma patients based on the seer. database 2021
8. Shu M, Bowen RS, Herrmann C, et al. Deep survival analysis with longitudinal X-rays for COVID-19. Proceedings of the IEEE/CVF Int Conference Computer Vision. 2021:4046-55.
9. https://www.kaggle.com/code/saychakra/survival-analysis-of-lung-cancer-

patients/data

10. ALPAR R. Applied multivariate statistical methods. 5th Edition Detay Publishing House, Ankara, 2017; 400.

11. Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society. 1972;34:187-202.

12. Ozdamar K. Biostatistics with SPSS. 4th Edition. Kaan Publishing House, Eskişehir, 2001.

13. Garcia FCC, Hirao A, Tajika A, et al. Leveraging longitudinal lifelog data using survival models for predicting risk of relapse among patients with depression in remission. Annual Int Conf IEEE Eng Med Biol Soc. 2021:2455-8.

14. Faraggi D, Simon R. A neural network model for survival data. Stat Med. 1995;14:73-82.

15. Kvamme, H., Borgan, Ø., & Scheel, I. Time-to-event prediction with neural networks and Cox regression. JMLR. 2019;1-30.

16. Englebert C, Quinn T, Bichindaritz I. Feature selection for survival analysis in bioinformatics. in proceedings of the workshop on advances in bioinformatics and artificial intelligence: bridging the gap co-located with 26th International Joint Conference on Artificial Intelligence (IJCAI). 2017;30-5.

17. Harrell Jr FE, Lee KL, Califf RM, et al. Regression modelling strategies for improved prognostic prediction. Stat Med. 1984;3:143-52.

18. Zhao L, Feng D. Deep neural networks for survival analysis using pseudo values. IEEE J Biomed Health Inform. 2020;24:3308-14.

19. Oei RW, Lyu Y, Ye L, et al. Progression-free survival prediction in patients with nasopharyngeal carcinoma after intensity-modulated radiotherapy: machine learning vs. traditional statistics. J Pers Med. 2021;11:787.

20. Lutz M. Programming python: O'Reilly Media, Inc; 2001.

21. Sertkaya D, Nihal A, Sozer MT. Cox regression model with time dependent covariate in survival anaysis. Journal of Ankara University Faculty Medicine. 2005;58:153-8.

22. Zhu X, Yao J, Huang J, et al. Deep convolutional neural network for survival analysis with pathological images. IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2016:544-7.