A Deep Inverse Weibull Network

Ola Abuelamayem [1,*]

[1]*Department of Statistics, Faculty of Economics and Political Science, Cairo University, Egypt*

Abstract   Survival analysis is heavily used in different field like economics, engineering and medicine. The main core of the analysis is to understand the relationship between the covariates and the survival function. The analysis can be performed using traditional statistical models or neural networks. Recently, neural networks has attracted attention in analyzing lifetime data due to its fl xibility in handling complex covariates. The networks introduced in the literature have some restrictions such as proportional hazard assumption, data discretization, monotonicity of hazard rates and heavy tailed assumption. In this paper, a novel neural network is introduced based on inverse Weibull distribution and random censoring that removes some of the restrictions introduced in the literature. The network doesn't put monotonicity, proportionality or heavy tailed assumptions on the hazard function. Also, the network doesn't require data discretization. To test its applicability, the network is applied on both simulated and real datasets and the numerical results show that our model outperforms some other methods in the literature.

Keywords   Inverse Weibull, Deep learning, Neural Network, Random Censoring.

AMS 2010 subject classifications 62M45; 62N01; 62N02; 62N05;

DOI: 10.19139/soic-2310-5070-2070

1. Introduction

Time to event analysis or survival analysis is an important approach in statistics that deals with lifetime data and different types of censoring. It has applications in different field like medicine, engineering, financ and economics. For example, time till cardiovascular death, time till cancer recurrence, time till failure of plain system. Usually, time to event is not observed for all data units producing so called censored data. If the lifetime of censored data is only known to occur after the recorded time, it is called right censoring.

The fundamental problem in analyzing right censoring time to event data is to study the relationship between different covariates and lifetime distribution. Inverse Weibull can be used as a lifetime distribution due to its fl xibility in handling different hazard shapes. In literature, inverse Weibull was heavily used to estimate right censored lifetime data. However, the analysis was too simple and doesn't take into consideration the effect of complex covariates and large data.

Analyzing data with different covariates can be done using traditional statistical models or neural networks. Recently, the latest has attracted attention due to its ability of handling complex covariates. Faraggi-Simon network (Faraggi and Simon (1995)) [4] was firs introduced as a nonlinear extension of Cox proportional hazard model. However, the network failed to outperform the traditional Cox model. Most modern Deep Learning (DL) survival models have been developed only since the late 2010s (Wiegrebe *et al* (2024)) [13]. Katzman *etal* (2016) [5]

---

*Correspondence to: Ola Abuelamayem (Email: ola.abuelamayem@feps.edu.eg). Department of Statistics, Faculty of Economics and Political Science, Cairo University, Egypt.

modifie Faraggi-Simon network by applying new deep learning techniques which outperforms the traditional Cox model. Zhu *et al* (2016) [14] and Zhu *et al* (2017) [15] introduced another extension for Cox model by analyzing images using a deep convolutional neural network. However, these studies have maintained the basic assumption of Cox model which is proportional hazard. Kvamme *et al* (2019) [7] solved this problem by introducing a non proportional Cox neural network. However, this model is still a relative risk one. Tong and Zhao (2022) [12] introduced an extension of Kvamme *et al* (2019) [7] network to impute missing data. A great review for survival neural networks can be found in Wiegrebe *et al* (2024) [13].

Lee *etal* (2018) [9] introduced DeepHit neural network that doesn't make any assumptions on the hazard function. However, the time is discretized based on a predefine maximum time horizon. Pawley (2020) [11] introduced a parametric neural network (DeepWeibull) based on Weibull distribution that allows to analyze continuous data. However, Weibull distribution can not analyze non monotone functions such as unimodal hazard function which is a restrictive assumption. In many real life situations, it is previously known that the hazard function cannot be monotone. For example, in a breast cancer study Langlands *et al* (1979) [8] observed that the peak mortality occurs after three years of surgery, and then it gradually decreases. Avati (2020) [3] introduced neural network using lognormal distribution. However, this network is limited to heavy tail charachteristic of the distribution which may lead to non accurate result if this is not the case. If the hazard function is non monotonic and there is no existance for heavy tail assumption, inverse Weibull can be used in the analysis.

In this paper, we introduce a novel neural network based on inverse Weibull loss function. This network doesn't assume a proportional hazard, relative risk, monotone hazards, data discretization or heavy tailed probability. The paper is organized as follows; In section 2, the basic definition for survival analysis and inverse Weibull distribution are introduced. In section 3, the network is introduced. In section 4, the analysis is performed on simulated and real datasets. Finally, the paper is concluded in section 5.

2. Inverse Weibull distribution and survival analysis

In this section, we introduce the basic definition of survival analysis and illustrate the usage of inverse Weibull as a lifetime distribution.

2.1. Survival analysis

Lifetime data is compromised of three elements: observed covariates (x), observed event time (T) and type of event indicator (δ) such that:

$$\delta = \left\{ \begin{array}{ll} 1, & \text{if the event is observed,} \\ 0, & \text{if the event is censored.} \end{array} \right. .$$

Survival and hazard are two main functions in time to event analysis. The survival function is denoted by $S(t) = P(T > t)$, which is the probability that an individual has survived up to time t. Hazard function $(h(t))$ is the instantaneous failure rate at time t, given that the individual has survived up to time t, it can be expressed mathematically as follows

$$h(t) = \lim_{\Delta t \to 0} \frac{p(\text{failure occurring in } [t, t + \Delta t | T > t))}{\Delta t} = \frac{f(t)}{S(t)},$$

where $f(t)$ is the probability density function.

2.2. Inverse Weibull distribution

From now, we assume that T follows inverse Weibull distribution. The probability density function, survival function and hazard function can be expressed respectively as follows

$$\begin{aligned}
f(t) &= \alpha\beta(\alpha t)^{-(\beta+1)}e^{-(\alpha t)^{-\beta}}, \\
S(t) &= 1 - e^{-(\alpha t)^{-\beta}}, \\
h(t) &= \frac{\alpha\beta(\alpha t)^{-(\beta+1)}e^{-(\alpha t)^{-\beta}}}{1 - e^{-(\alpha t)^{-\beta}}}.
\end{aligned}$$

Different shapes of $f(t)$ and $h(t)$ are illustrated in Figures 1 and 2, respectively. From Figure 1, It can be seen that α is the scale parameter and β is the shape parameter. Also, it is clear that inverse Weibull distribution takes different shapes according to the values of β. For example when $\beta < 1$ the *pdf* is monotonically decreasing, while when $\beta > 1$ the *pdf* increases till it reaches its peak and then monotonically decreases. The hazard rate can be monotonically decreasing, non monotonic, constant for short time then starts to increase till it reaches its peak and decreases, as illustrated in Figure 2.

In Figure 3, the hazard functions of inverse Weibull and log normal distributions are illustrated. It can be seen that, inverse Weibull relaxes the assumption of heavy tailed proposes in log normal network.

The likelihood function for n data points under right random censoring can be expressed as follows

$$L_1 = \prod_{i=1}^{n}[f(t_i)]^{\delta_i}[S(t_i)]^{1-\delta_i}.$$

Substituting by $f(t)$ and $S(t)$ for inverse Weibull distribution in the previous equation, the likelihood function is

$$L_1 = \prod_{i=1}^{n}[\alpha\beta(\alpha t_i)^{-(\beta+1)}e^{-(\alpha t_i)^{-\beta}}]^{\delta_i}[1 - e^{-(\alpha t_i)^{-\beta}}]^{1-\delta_i}. \tag{1}$$

Inverse Weibull distribution has several applications in different fields For example in health studies using the ability of inverse Weibull in analyzing non monotone hazards. For example, mortality of breast cancer has a non monotonic hazard function where the peak is after three years of surgery (Langlands *et al* (1979)) [8]. In engineering, Nelson (1982) [10] mentioned that inverse Weibull distribution provides a good fi to time to breakdown of an insulating flui subjected to the action of a constant tension data. Also, Keller *et al* (1983) [6] illustrated that inverse Weibull distribution significantl fi the failure data of dynamic diesel engine components better than the Weibull distribution.
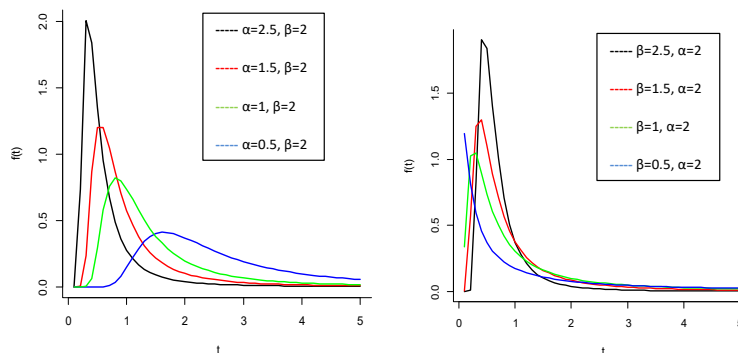

Figure 1: The probability density function for inverse Weibull distribution.
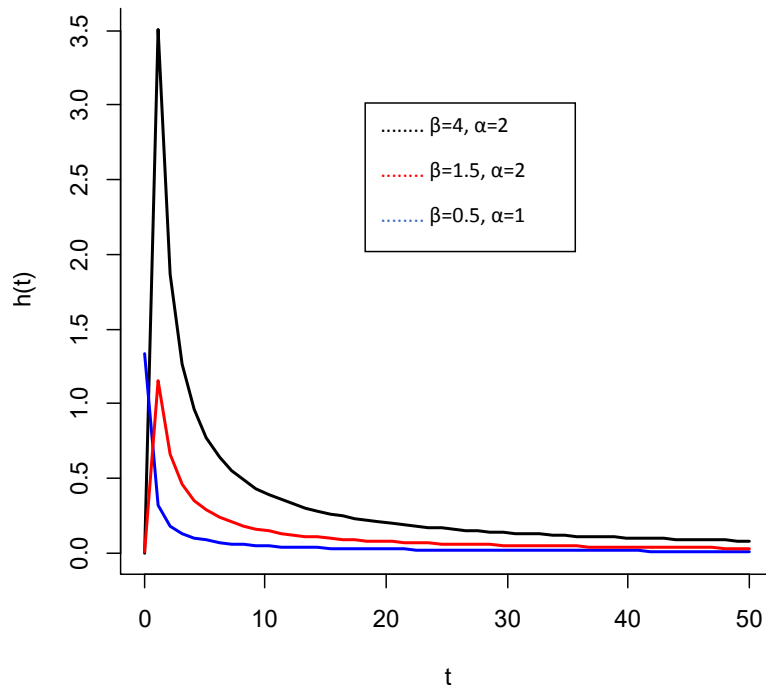
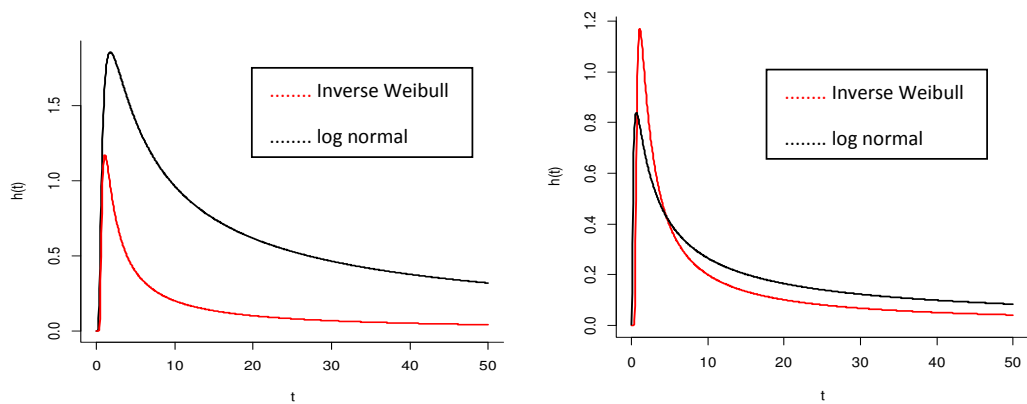Figure 2: The hazard function for inverse Weibull distribution

Figure 3: The hazard function for inverse Weibull vs log normal distributions

3. Model description

In this section, a brief introduction about neural networks is introduced. Then, deep inverse Weibull network is illustrated.

3.1. Neural networks

Neural networks is a group of interconnected units called neurons. Its main goal is to predict an output based on a functional relationship with the inputs. This function is called an activation function and can be illustrated mathematically as follows

$$z = f(b + \sum (w_i x_i))$$

Where:

b is called a bias term,

$x_1, x_2, ..., x_n$: a set of inputs

$w_1, w_2, ..., w_n$: weights

The choice of the activation function depends on the required range of the output. For example, if the output is a value between 0 and 1, sigmoid function could be a choice.

A second important component of the network is called loss function. It's a function to reflec the prediction error. Mean square error is a common choice for the loss function in most of the applications. However, in survival analysis one needs to take censoring information into consideration. This is done through the likelihood function. So, the loss function in this case is negative log likelihood function. The network updates the weights by minimizing the loss function, this can be done using backpropagation algorithm. For more details, see(Aggarwal (2018)) [1].

3.2. Deep inverse Weibull network

The goal of this model is to predict the parameters of inverse Weibull distribution using the neural network ability to handle complex covariates and maintaining the continuity of the data without proportional, monotonic hazard or heavy tailed assumptions. As illustrated in Figure 4, DeepInvWei is a deep neural network that takes as input the covariates and output the estimates of inverse Weibull distribution's parameters that fully characterize the survival function. To determine the network structure, different combinations were tried (different number of layers widths and activation functions) and the one with best performance is selected and has the following form

The hidden layers of the network consist of a fully connected layer followed by a dropout layer. It consists of f ve fully connected hidden layers of widths 1, 6, 4, 2 and 1 of the covariates dimension with softmax and tanh activations, respectively. The output layer has a softplus activation function to consist with the parameters' range $]0,\infty[$.

For training, we use back-propagation via the Adam optimizer. Each of the hidden layers have Xavier initialization and a dropout rate of 0.1 or 0.2. DeepInvWei is implemented in a TensorFlow environment with the Keras API.

The loss function is define to be the negative log likelihood for random right censored inverse Weibull data (eq. 1), which can be written as follows

$$L = -\sum_{i=1}^{n} (\delta_i) \ln[\alpha_i \beta_i (\alpha_i t_i)^{-(\beta_i+1)} e^{-(\alpha_i t_i)^{-\beta_i}}] + (1 - \delta_i) \ln[1 - e^{-(\alpha_i t_i)^{-\beta_i}}]. \tag{2}$$

In Figure 5, we illustrate a computational graph to compute the training loss of the proposed network: the inputs are the covariates x and the outputs are α and β.

output layer

α, β

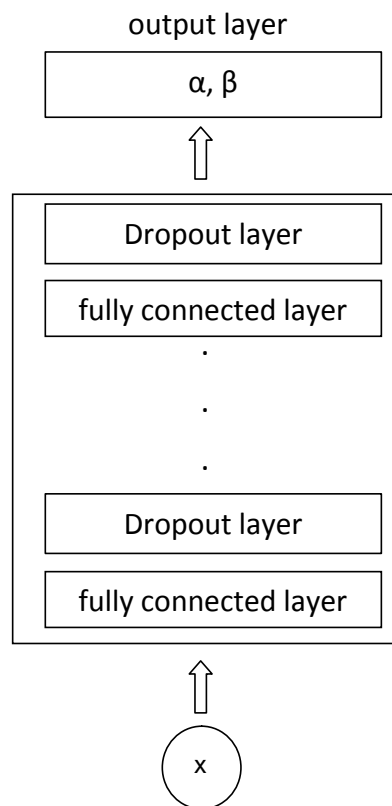| Dropout layer |
|---|
| fully connected layer |
| . |
| . |
| . |
| Dropout layer |
| fully connected layer |

x

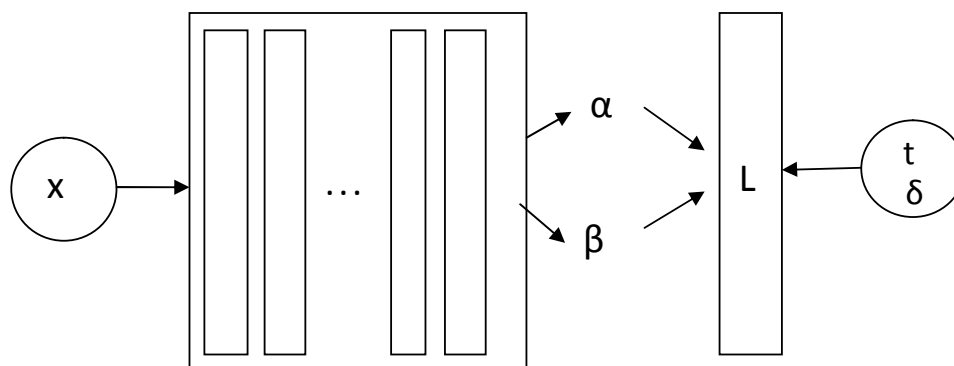Figure 4: The architecture of DeepInvWei.

α

β

...

L

t
δ

x

Figure 5: Computational graph to compute the training loss of DeepInvWei

4. Experiments

Eight sets of experiments based on synthetic datasets and four sets of experiments based on real datasets are analyzed. In Table 1, an overview on some descriptive statistics of the four real datasets is given. To evaluate the performance of DeepInvWei, we compare the results with the most widely used network in the medical fiel as it outperforms other methods in the literature (DeepHit). Also, we compare it with the parametric neural network DeepWeibull. For log normal neural network, the structure of the network is not clearly illustrated (i.e. the optimum number of layers and width) which is essential for network application. Accordingly, Only DeepWeibull as a parametric model was used in the comparison.

Table 1: Descriptive statistics of real datasets.

		No. of uncensored	No. of censored	No. of features
Datasets	METABRIC	1103	801	9
	SUPPORT	6036	2810	14
	FLCHAIN	1962	4562	8
	NWTCO	668	3360	6

Molecular Taxonomy of Breast Cancer International Consortium (METABRIC)

METABRIC dataset consists of gene expression features and clinical data for 1904 patients. It aims to help physicians classifying new cancer subgroups. Similar to Katzman *et al* (2018) [5], we used in the analysis four gene indicators (MKI67, EGFR, PGR, and ERBB2) and f ve clinical features (hormone treatment indicator, radiotherapy indicator, chemotherapy indicator, ER-positive indicator, age at diagnosis). The data is obtained from DeepSurv python package. 1103 patients were followed up until death and the remaining 801 patients were right censored.

Study to Understand Prognoses Preferences Outcomes and Risks of Treatment (SUPPORT)

SUPPORT dataset consists of 14 different features for 8873 patients. It aims to study the survival time of seriously ill hospitalized adults. We used in the analysis 14 features (age, sex, race, number of comorbidities, presence of diabetes, presence of dementia, presence of cancer, mean arterial blood pressure, heart rate, respiration rate, temperature, white blood cell count, serums sodium, and serums creatinine). The data is obtained from DeepSurv python package. 6036 patients were followed up until death and the remaining 2810 patients were right censored.

Assay Of Serum Free Light Chain (FLCHAIN)

FLCHAIN dataset consists of 9 different covariates for 6524 patients. It aims to study the relationship between serum free light chain (FLC) and mortality. Similar to Kvamme *et al* (2019) [7] we discarded the chapter covariate and use in the analysis the remaining eight features (age, sex, the calendar year in which a blood sample was obtained, kappa portion, lambda portion, the FLC group for the subject, serum creatinine, monoclonal gammapothy indicator). The data is obtained from DeepSurv python package. 1962 patients were followed up until death and the remaining 4562 patients were right censored.

Data from the National Wilm's Tumor (NWTCO)

NWTCO dataset consists of 6 different covariates for 4028 patients. It aims to predict the survival based on tumor histology. We used in the analysis 6 features (Histology from local institution, Histology from central lab, Disease stage, study, age, sub-cohort). The data is obtained from DeepSurv python package. 668 patients were followed up until death and the remaining 3360 patients were right censored.

Synthetic

We generated eight synthetic datasets with the following characteristics:

Large linear Inverse Weibull (LLIW): Inverse Weibull model with linear relationships between the covariates and the parameters with 40,000 sample units.

Small linear Inverse Weibull (SLIW): Inverse Weibull model with linear relationships between the covariates and the parameters with 1000 sample units.

Large non-linear Inverse Weibull (LNIW: Inverse Weibull model with quadratic relationships between the covariates and the parameters with 40,000 sample units.

Small non-linear Inverse Weibull (SNIW: Inverse Weibull model with quadratic relationships between the covariates and the parameters with 1000 sample units.

The data is generated as follows

1) For each unit ($i = 1, ..., n$):

  a) generate two covariates from uniform(-1,1)

  b) Set

$$\begin{aligned} \alpha_i &= 1 + 2x_{1i}, & \text{for linear data} \\ \alpha_i &= x_{1i}^2 + 2x_{2i}, & \text{for non-linear data} \end{aligned}$$

  c) Generate t_i from inverse Weibull with parameters α_i and ($\beta = 1$, $\beta = 3$).

2) Randomly select 20% of the units to be censored.

3) For censored data, generate a right censored event time from uniform $(0, t_i)$.

4) Set $\delta_i = 0$ for censored data, otherwise $\delta_i = 1$.

5) If $t_i > 500$, set $t_i = 500$. This is done to keep the harmony of the data.

Model evaluation

For evaluation, the data is divided into train/test datasets. However, performing this process once may lead to misleading prediction error. For more robust result, k fold cross validation is applied. The mean and standard deviation among k fold cross validations is presented.

The data is divided into three randomly train/test datasets, with 80% of the units in each training set and the remaining 20% in the test set. We reserved 20% of the training set as a validation set. All datasets are generated with 20% censoring. To test the performance of the model, we used time dependent concordance index (C^{td}) (Antolini *et al*(2005) [2]) which has the following formula

$$C^{td} = Pr\{S(T_i|X_i(t)) < S(T_j|X_j(t))|T_i < T_j, \delta_i = 1\}$$

The concordance index for synthetic and real datasets is illustrated in Tables 2 till 4, respectively. High C^{td} indicates better model performance in learning the patients' survival distribution.

In synthetic datasets we can obtain the oracle C^{td} using the true survival distribution. This helps in providing a benchmark against which to compare the model. If the model approaches near the oracle then it has likely learned the true distribution. Sometimes due to random chance, a model may beat the oracle metric (Pawley (2020)) [11]. From Tables 2 and 3, It can be seen that the concordance index for DeepInvWei is very close to the oracle one, which indicates the model's ability to learn the true survival distribution. Also, It is clear that DeepInvWei has the highest C^{td} which is expected as the data has a non monotonic hazard function and is generated from inverse Weibull distribution. Also, it is illustrated that for data with higher values of hazard function ($\beta = 3$), DeepInvWei outperforms other networks with a big difference.

For the standard deviation in synthetic data, it seems that DeepHit outperforms others in majority of the cases. However, DeepInvWei still gives a relatively small standard deviation. In terms of its higher discrimination power, it is suggested to use DeepInvWei.

In real world, we can't guarantee that the data is generated from inverse Weibull distribution. To compare the model and test its applicability, we apply it on four different real datasets. From table 4, it can be seen that, DeepInvWei has the highest C^{td} in almost all datasets. It provides performance improvements over other models; with the exception of DeepHit in METABRIC dataset. In FlCHAIN and NWTCO, DeepInvWei has smaller standard deviation than others.

Table 2: Concordance Index (standard deviation) for synthetic datasets ($\beta = 1$).

		Oracle	DeepHit ($\alpha = 0$)	DeepWeibull	DeepInvWei
	LLIW	0.5692	0.4640(0.0135)	0.5348(0.0138)	**0.5729**(0.0250)
	SLIW	0.5650	0.5444(0.0258)	0.5456(0.0284)	**0.6447**(0.0290)
Datasets	LNIW	0.6783	0.4726(0.0090)	0.5485(0.0456)	**0.6703**(0.0286)
	SNIW	0.6645	0.5748(0.0420)	0.5691(0.0876)	**0.6422**(0.0473)

Table 3: Concordance Index for synthetic datasets ($\beta = 3$).

		Oracle	DeepHit ($\alpha = 0$)	DeepWeibull	DeepInvWei
	LLIW	0.6159	0.4285(0.0076)	0.4686(0.0598)	**0.6529**(0.0490)
	SLIW	0.6024	0.3873(0.0009)	0.4771(0.0347)	**0.6530**(0.0919)
Datasets	LNIW	0.6394	0.4260(0.0040)	0.5307(0.0648)	**0.6386**(0.1612)
	SNIW	0.6192	0.5267(0.1181)	0.5478(0.0587)	**0.6289**(0.0527)

Table 4: Concordance Index (standard deviation) for real datasets.

		DeepHit ($\alpha = 0$)	DeepWeibull	DeepInvWei
	METABRIC	**0.6223**(0.0080)	0.5841(0.0189)	0.6166(0.0108)
	SUPPORT	0.5455(0.0131)	0.5515(0.0055)	**0.5586**(0.0083)
Datasets	FLCHAIN	0.7378(0.0128)	0.6102(0.0071)	**0.7661**(0.0055)
	NWTCO	0.5194(0.024)	0.5338(0.0466)	**0.6977**(0.0087)

5. Limitations and future Work

A direct generalization to this network can be done by changing the family of distribution used. Also, backpropagation is used in neural network training, other algorithms could be used. Moreover, more search could be done to reduce the variability of the network. Finally, one can apply Bayesian network to get use of any pre-existing information.

6. Conclusion

This paper presents a new model (DeepInvWei) to be used in the survival analysis. It trains a neural network to predict the parameters of inverse Weibull distribution with respect to the covariates while capturing the right censoring nature. DeepInvWei has the fl xibility of analyzing continuous data without discretization. Also, it has the advantage of using inverse Weibull properties. For example, it can analyze data with non-monotonic hazards unlike Weibull distribution. In many real life situations, it is previously known that the hazard function cannot be monotone. For example, in a breast cancer study it is known that the peak mortality occurs after three years of surgery, and then it gradually decreases. As a test, we compared the performance of DeepInvWei with the performance of previous models. The numerical results showed that, for majority of cases, DeepInvWei performed better.

REFERENCES

1. Aggarwal, C. *Neural Networks and Deep Learning*, Springer nature, Springer Nature Switzerland AG, 2nd ed., 2023. DOI https://doi.org/10.1007/978-3-031-29642-0

2. Antolini, L., Boracchi, P. and Biganzoli, E, *A time-dependent discrimination index for survival data*, Statistics in Medicine, vol. 24, pp. 39273944, 2005.

3. Avati, A., Duan, T., Jung, K., Shah, N. H., and Ng, *Countdown Regression: Sharp and Calibrated Survival Predictions*, Proceedings of The 35th Uncertainty in Artificia Intelligence Conference, PMLR, vol. 115, pp. 739–749, 2020.

4. Faraggi, D., and Simon, R, *A neural network model for survival data*, Statistics in Medicine, vol. 14, pp. 73–82, 1995.

5. Katzman, J., Shaham, U., Cloninger, A., Bates, J., Jiang, T. and Kluger,Y, *DeepSurv: Personalized treatment recommender system using a cox proportional hazards deep neural network.*, arXiv:1606.00931, 2016, [online] Available: http://arxiv.org/abs/1606.00931.

6. Keller, A., Giblin, M. and Farnworth, N, *Reliability Analysis of Commercial Vehicle Engines*, Reliability Engineering, vol. 10, no. 1, pp. 15–25, 1983.

7. Kvamme, H., Borgan, Borgan, O. and Scheel, I, *Time-to-Event Prediction with Neural Networks and Cox Regression*, Journal of Machine Learning Research, vol. 20, pp. 1–30, 2019.

8. Langlands, A., Pocock, S., Keer, G. and Gore, S, *Long-term survival of patients with breast cancer: a study of the curability of the disease*, British MedicalJournal, vol. 2, no.6200 , pp. 1247–1251, 1979.

9. Lee, C., Zame, W., Yoon, J. and Schaar, M, *DeepHit: A Deep Learning Approach to Survival Analysis with Competing Risks*, Proceedings of the thirty-second AAAI conference on artificia intelligence, pp. 2314-2321, 2018.

10. Nelson W, *Applied Life Data Analysis*, John Wiley andSons, USA.DOI: https://doi .org/10.1002/0471725234, pp. 739–749, 1982.

11. Pawley, M, *eepWeibull: a deep learning approach to parametric survival analysis*, M.Sc. Thesis, Departement of Mathematics, Imperial College London, 2020.

12. Tong, J. and Zhao, X, *Deep survival algorithm based on nuclear norm*, Journal of statistical computation and simulation, vol. 92, no. 9, pp. 1964–1976, 2022.

13. Wiegrebe, S., Kopper, P., Sonabend, R. Bischl, B. and Bender, A, *Deep learning for survival analysis: a review*, Artificia Intelligence Review, vol. 57, no. 65, 2024.https://doi.org/10.1007/s10462-023-10681-3.

14. Zhu,X., Yao, J., Zhu, F. and Huang, H, *WSISA: Making Survival Prediction from Whole Slide Histopathological Images*, IEEE Conference on Computer Vision and Pattern Recognition, 2017. DOI: 10.1109/CVPR.2017.725.

15. Zhu, X., Yao, J and Huang, J, *Deep Convolutional Neural Network for Survival Analysis with Pathological Images*, IEEE International Conference on Bioinformatics and Biomedicine. DOI: 10.1109/BIBM.2016.7822579.