Journal of Innovations in Computer Science and Engineering
Shahid Beheshti Unversity

June 2025, Volume 3, Issue 1

Deep Learning Frailty Model for Heart Failure Survival Prediction

Solmaz Norouzi✉, **Code Orcide: 0000-0001-5805-6072**
Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran.
snorouzibiostatistics@gmail.com
Ebrahim Hajizadeh, **Code Orcide: 0000-0001-7863-4837**
Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran.hajizadeh@modares.ac.ir
Mohammad Asghari Jafarabadi, **Code Orcide: 0000-0003-3284-9749**
Biostatistics Unit, School of Public Health and Preventive Medicine, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, VIC, 3004, Australia.
Department of Psychiatry, School of Clinical Sciences, Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, VIC, 3168, Australia. m.asghari862@gmail.com
Hossein Khormaei, **Code Orcide: 0009-0004-3274-5776**
Department of Electrical Engineering, National University of Skills (NUS), Tehran, Iran. hosainkhormaee@gmail.com
Nasim Naderi, **Code Orcide: 0009-0000-7241-7460**
Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran.
naderi.nasim@gmail.com

Abstract—The study employed Deep Learning Frailty (DLF), a compelling neural modeling framework for predicting heart failure patient survival. The DLF embeds a notion of multiplicative frailty from classical survival analysis that deals with unobserved heterogeneity while exploiting the neural structure's strong capabilities in approximating any non-linear covariate relationship. The results showed that Incorporating frailty leads to significant improvements, and the DLF model performs better on average.

Keywords—Deep Learning, prediction, survival analysis, heart failure

I. INTRODUCTION

The survival analysis of patients is an important section in medical research, as it estimates the time to specific events(1). Deep learning methods have recently become quite popular for modeling complex relationships within survival data. In this work, the authors investigate advanced applications of deep neural networks in survival analysis within the framework of frailty models. Frailty is the unobserved heterogeneity that influences the occurrence of the event. Combining the ideas could, therefore, enhance the accuracy and interpretability of survival predictions in domains with multiple possible outcomes(2, 3).
Traditional survival analysis models usually assume that the data are independently and identically distributed, which might be the reason for biased estimates, hence reducing their accuracy(4). Therefore, sophisticated methods are required to capture nonlinear relationships within such models. This paper proposes a new method of incorporating frailty models into deep neural networks to perform survival analysis on heart failure patients by considering the dependency structure among the observations explicitly. The present study aims to develop a more powerful framework for predicting survival outcomes and compare its performance with DeepSurv(5, 6).

II. LITERATURE REVIEW

Recent advancements in deep learning have significantly impacted survival analysis, particularly in addressing the challenges posed by unobserved heterogeneity. Lee et al. (2018) introduced DeepHit, a non-parametric model that surpasses traditional methods by learning the distribution of survival times without restrictive parametric assumptions (7). To mitigate the negative impact of irrelevant features, Rietschel et al. (2018) proposed innovative feature selection techniques that significantly improved deep learning model performance in medical datasets (8). Huang and Liu (2020) developed DeepCompete. This continuous-time model effectively handles competing risks (9), while Nagpal et al. (2020) presented Deep Survival Machines. This fully parametric model outperforms traditional methods in handling censored data without relying on the Cox proportional hazards assumption (10).

Recent studies have explored the integration of frailty models into deep learning frameworks to address unobserved heterogeneity (11). Tran et al. (2020) and Mendel et al. (2022) incorporated random effects into deep neural networks to account for frailty (۱۳ ,۱۲), while Hangbin Lee et al. (2023) proposed the DNN-FM model, which effectively handles censored survival data and improves predictive performance (۱٤). Wu et al. (2023) introduced the Neural Frailty Machine (NFM), which incorporates multiplicative frailty to address unobserved heterogeneity and demonstrates superior predictive capabilities across various datasets (۱٥).

Submit Date: 2025-03-22
Accept Date: 2025-09-28
✉ Corresponding author

FRAILTY MODELS, A CORNERSTONE OF MODERN SURVIVAL ANALYSIS, EXTEND THE COX MODEL BY INTRODUCING A MULTIPLICATIVE RANDOM EFFECT TO ACCOUNT FOR UNOBSERVED HETEROGENEITY (١٦, ١٧)

While the theoretical foundations of frailty models are well-established, most research has focused on linear relationships (١٨, ١٩, ٢٠).

By leveraging deep learning, frailty models can capture complex, non-linear relationships and address censoring and competing risks through the frailty parameter. This flexible approach accurately estimates unobserved heterogeneity and correlations between different event types. Additionally, it improves the precision of covariate effect estimates based on specific event causes, ultimately enhancing the model's accuracy and predictive performance.

III. MATERIALS AND METHODS

A. Study Population

Data from 529 patients admitted into the RCMCH information system between March and August 2018 were first retrieved for analysis. They were diagnosed with HFrEF and received standard treatment. After the exclusion of those patients who did not undergo standard treatment, the data from 435 patients will be included in the survival time analysis, followed from 2018 to July 2023, equivalent to 5 years. This dataset consists of 57 demographic and clinical features.

B. Statistical Analysis

All the raw data in this investigation were put together in the form of a database, and further tests were run on the processed data. These analyses have been implemented through Python using Pytorch by implementing both DLF and DeepSurv models. There were standard training/validation splits in the survival dataset, so a 5-fold cross-validation was done with one fold reserved for testing and 20% used for validation. Model performance was evaluated using three metrics standard in survival predictions, namely the integrated Brier score (IBS), integrated negative binomial log-likelihood (INBLL), and c-index.

DLF considers different deep neural network architectures for the modeling of survival analysis. The central insight driving this work is the way censored observations enter into a likelihood to produce consistent parameter estimates, given partial information about event times. Extension to include frailty into the deep neural network model enables the DLF structure to capture individual-specific traits or temporal changes driven by some unobserved factors that possibly influence event risk. Intra-individual correlation can be induced by unobservable individual-specific factors.

For the frailty variable u, we utilize the gamma density function;

$$G_\theta(x) = \frac{1}{\theta}\log(1+\theta x)\,, \theta \geq 0 \qquad (1)$$

We begin by integrating the conditional survival function with frailty to derive the observed likelihood function:

$$\begin{aligned} S(t \mid X) &= \mathbb{E}_{u_i \sim f_\theta}\left[e^{-u_i \int_0^t e^{h(s,X)} ds}\right] \\ &=: e^{-G_\theta\left(\int_0^t e^{h(s,X)} ds\right)} \end{aligned} \qquad (2)$$

The frailty transform ($G_\theta(x) = -\log\left(\mathbb{E}_{u_i \sim f_\theta}[e^{-u_i x}]\right)$ is defined as the $(-\log)$ Of the Laplace transform of the frailty distribution for each cause. Consequently, the conditional cumulative hazard function is given by $\mathrm{H}(t|X) = G_\theta\left(\int_0^t e^{h(s,X)} ds\right)$.

For the PF model, we utilize two Multi-Layer Perceptrons (MLPs), denoted as $\hat{h} = \hat{h}(t; W^h, b^h)$ and $\hat{m} = \hat{m}(X; W^m, b^m)$, to approximate the functions h and m, parameterized by (W^h, b^h) and (W^m, b^m), respectively. Here, W represents a collection of weight matrices across all layers of the MLPs, while b denotes a set of bias vectors across all layers. Considering the standard results regarding the likelihood of censored data as presented in Equation (2), the learning of parameters under the PF framework can be expressed as follows (26).

$$\begin{aligned} &\mathcal{L}(\mathrm{W}^h, \mathrm{b}^h, \mathrm{W}^m, \mathrm{b}^m, \theta) \\ &= \frac{1}{n}\left[\sum_{i\in[n]} \delta_i \log g_\theta\left(e^{\hat{m}(X_i)} \int_0^{T_i} e^{\hat{h}(s)} ds\right) + \delta_i \hat{h}(T_i) + \delta_i \hat{m}(X_i) \right. \\ &\qquad \left. - G_\theta\left(e^{\hat{m}(X_i)} \int_0^{T_i} e^{\hat{h}(s)} ds\right)\right] \end{aligned} \qquad (3)$$

in which $g_\theta(x) = \frac{\partial}{\partial x} G_\theta(x)$.

The estimated conditional cumulative risk and survival functions are represented by equation (4).

$$\hat{\mathrm{H}}_{\mathrm{DLF}}(t \mid X) = G_{\hat{\theta}_n}\left(\int_0^t e^{\hat{h}_n(s)+\hat{m}_n(X)} ds\right), \tag{4}$$
$$\hat{S}_{\mathrm{DLF}}(t \mid X) = e^{-\hat{\mathrm{H}}_{\mathrm{DLF}}(t \mid X)},$$

IV. Results

This study analyzed the mortality of 435 heart failure patients over five years, focusing on deaths from heart failure. At the study's conclusion, 29.4% of patients were alive, and 40.9% were censored. Of the patients, 43.96% died from heart failure. The median survival time was 43.40 months.

The one-year survival rate for patients who died from heart failure was 80.66% (95% CI: 0.76-0.84), decreasing to 68.3% (95% CI: 0.63-0.72) at three years and 59.52% (95% CI: 0.54-0.64) at five years. For those who died from other causes, the survival rates were 91.78% (95% CI: 0.88-0.94) at one year, 79.08% (95% CI: 0.74-0.83) at three years, and 70.29% (95% CI: 0.64-0.75) at five years.

The mean age of all patients was 18.11 ± 56.57 years, ranging from 14 to 95. Among those who died from heart failure, the average age was 59.26 ± 1.40 years, with the highest mortality in the 56-65 age group. Of these, 63.1% were male, 89.4% self-employed, and 87.5% held a bachelor's degree. Additionally, 93.1% resided in urban areas, and 89.4% were married.

To identify the optimal learning rate for the deepsurv and DLF models, we conducted experiments using a range of learning rates, specifically [0.1, 0.01, 0.001, 0.0001, 1e-05], and evaluated performance based on the c-index criterion. This process involved creating and deploying a new deepsurv model for each learning rate, utilizing a Negative Loglikelihood loss function and the Adam optimizer over 50 epochs, during which learning rate and weight decay were monitored. We systematically compared model performance across these learning rates and determined the optimal rate to be lr = 0.006 (Fig 1).

We assessed the performance gains of DLF relative to their non-frailty counterpart, DeepSurv. As shown in Table 1, incorporating frailty leads to significant improvements, with performance increases exceeding 10% for DLF models in IBS and INBLL metrics. This enhancement is consistent across the HF dataset. The results suggest that the DLF model performs better on average.

TABLE I. DLF MODELS IN COMPARISON TO THE NON-FRAILTY DEEPSURV MODEL.

Model	**metrics**		
	IBS	***INBLL***	***c-index***
DeepSurve	0.27±0.02	0.62±0.01	0.55±0.04
DLF	0.17±0.01	0.53±0.02	0.66±0.04

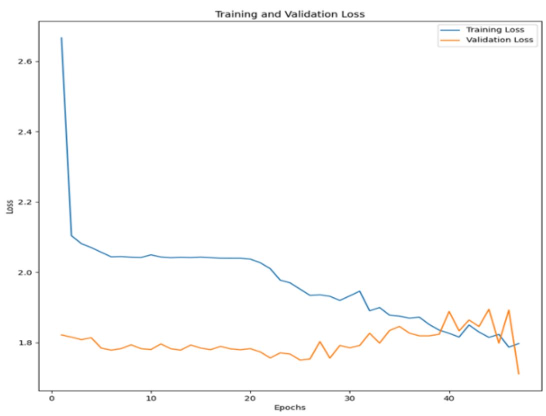

Fig 1: Assessing model performance on both training and validation datasets

V. Discussion

This study aimed to evaluate the accuracy of a deep learning model incorporating a frailty approach applied to heart failure data. The Deep Neural Frailty(DLF) method introduces an innovative approach to modeling mortality by utilizing two distinct neural network structures.

The results demonstrate that the DLF framework effectively handles censored data in survival analysis, leading to improved predictive accuracy, reduced bias, enhanced robustness, and more personalized predictions. These advantages make it a valuable tool for healthcare applications, where censored data is common. Our findings also indicate that incorporating frailty into the deep learning model significantly improves its predictive accuracy when applied to heart failure data. Frailty, in the context of time-to-event modeling, accounts for unobserved heterogeneity among individuals, which can influence the occurrence of an event. Frailty models extend traditional survival models, such as the Cox model, by introducing random effects that capture individual-specific factors not directly observed but influencing the event of interest. By including frailty,

time-to-event models better account for variability in event times that cannot be explained by the measured covariates alone. This study also compared two survival models: The DeepSurv model, and the DLF model with proportional frailty. Despite the small sample size, results showed that both the DeepSurv and DLF models outperform others in terms of accuracy in predicting survival time, thanks to their advanced ability to model complex relationships. This comparison helps guide researchers in selecting the most suitable model for survival analysis. These models excel at capturing complex, non-linear relationships between input variables and survival outcomes. Supporting our results, a study by Ruofan Wu et al. examined the performance of the Neural Frailty Machine (NFM) on survival data from five datasets, demonstrating that NFM outperformed 12 other reference models, particularly on the METABRIC, SUPPORT, and MIMIC-III datasets. NFM showed significant improvements in evaluation metrics such as IBS and INBLL compared to other models (15).
However, some studies argue that there is no conclusive evidence to suggest that deep learning models consistently outperform classical models. While deep learning is recognized for its ability to model complex and non-linear relationships, especially in survival time prediction, its performance is not always significantly superior to that of traditional models. Classical models can sometimes perform just as well, or even better, in certain cases. Additionally, some studies suggest that deep learning models require larger datasets for training and fine-tuning and may not consistently outperform traditional models when data is scarce (21). Both the DLF and the classic Cox model are important tools for analyzing competing risks, but they differ in structure and application. Ultimately, the choice of model depends on the specific characteristics of the data and the research context, meaning that deep learning models are not universally superior to classical models (۲۳ ,۲۲).

References

1. Kleinbaum DG, Klein M. Survival analysis a self-learning text: Springer; 1996.
2. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association. 1958;53(282):457-81.
3. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. Circulation. 2016;133(6):601-9.
4. Austin PC, Fine JP. Accounting for competing risks in randomized controlled trials: a review and recommendations for improvement. Statistics in medicine. 2017;36(8):1203-9.
5. Nagpal C, Li X, Dubrawski A. Deep survival machines: Fully parametric survival regression and representation learning for censored data with competing risks. IEEE Journal of Biomedical and Health Informatics. 2021;25(8):3163-75.
6. Zhong Q, Mueller JW, Wang J-L. Deep extended hazard models for survival analysis. Advances in Neural Information Processing Systems. 2021;34:15111-24.
7. Lee C, Zame W, Yoon J, Van Der Schaar M, editors. Deephit: A deep learning approach to survival analysis with competing risks. Proceedings of the AAAI conference on artificial intelligence; 2018.
8. Rietschel C, Yoon J, van der Schaar M. Feature selection for survival analysis with competing risks using deep learning. arXiv preprint arXiv:181109317. 2018.
9. Huang P, Liu Y, editors. Deepcompete: A deep learning approach to competing risks in continuous time domain. AMIA annual symposium proceedings; 2020: American Medical Informatics Association.
10. Nagpal C, Yadlowsky S, Rostamzadeh N, Heller K, editors. Deep Cox mixtures for survival regression. Machine Learning for Healthcare Conference; 2021: PMLR.
11. Hong C, Yi F, Huang Z. Deep-CSA: Deep Contrastive Learning for Dynamic Survival Analysis With Competing Risks. IEEE Journal of Biomedical and Health Informatics. 2022;26(8):4248-57.
12. Mandel F, Ghosh RP, Barnett I. Neural networks for clustered and longitudinal data using mixed effects models. Biometrics. 2023;79(2):711-21.
13. Tran M-N, Nguyen N, Nott D, Kohn R. Bayesian deep net GLM and GLMM. Journal of Computational and Graphical Statistics. 2020;29(1):97-113.
14. Lee H, HA I, Lee Y. Deep Neural Networks for Semiparametric Frailty Models via H-likelihood. arXiv preprint arXiv:230706581. 2023.
15. Wu R, Qiao J, Wu M, Yu W, Zheng M, Liu T, et al. Neural Frailty Machine: Beyond proportional hazard assumption in neural survival regressions. Advances in Neural Information Processing Systems. 2023;36:5569-97.
16. Wienke A. Frailty models in survival analysis: Chapman and Hall/CRC; 2010.
17. Duchateau L, Janssen P. The frailty model: Springer; 2008.
18. Murphy SA. Consistency in a proportional hazards model incorporating a random effect. The Annals of Statistics. 1994;22(2):712-31.
19. Parner E. Asymptotic theory for the correlated gamma-frailty model. The Annals of Statistics. 1998;26(1):183-214.
20. Kosorok MR, Lee BL, Fine JP. Robust inference for univariate proportional hazards frailty regression models. 2004.
21. Jung W, Kim SE, Kim JP, Jang H, Park CJ, Kim HJ, et al. Deep learning model for individualized trajectory prediction of clinical outcomes in mild cognitive impairment. Frontiers in Aging Neuroscience. 2024;16:1356745.
22. Wang M, Greenberg M, Forkert ND, Chekouo T, Afriyie G, Ismail Z, et al. Dementia risk prediction in individuals with mild cognitive impairment: a comparison of Cox regression and machine learning models. BMC Medical Research Methodology. 2022;22(1):284.
23. Kantidakis G, Biganzoli E, Putter H, Fiocco M. A simulation study to compare the predictive performance of survival neural networks with Cox models for clinical trial data. Computational and Mathematical Methods in Medicine. 2021;2021(1):2160322.