MALAYSIAN JOURNAL OF SCIENCE
MJS
ISSN 1394-3065

Regular Issue

Malaysian Journal of Science 41 (3): 44-48 (October 2022) https://mjs.um.edu.my

MORTALITY PREDICTION OF SURGICAL INTENSIVE CARE UNIT PATIENTS USING DEEP LEARNING-BASED SURVIVAL MODELS

Lintu M.K.[1a], S. N. S. Acharya[2b]*, Asha Kamath[3a], and David Raj Micheal[4c]

Abstract: Mortality prediction in surgical intensive care units (SICUs) is considered to be among the most critical steps in enforcing efficient treatment policies. This study aims to evaluate the performance of various deep learning models in predicting the mortality of patients admitted to SICUs. The survival of 2,225 adult patients admitted to SICUs was modeled using five salient deep learning-based survival models, namely, Cox-CC, Cox-Time, DeepSurv, DeepHit, and N-MTLR. The data were extracted from the Medical Information Mart for Intensive Care II (MIMIC-II) database. The performance of the models was compared using the time-dependent concordance index (C^{td}-index) and integrated Brier score (IBS). From among the five models, DeepSurv achieved the most accurate prediction, while Cox-Time demonstrated the least optimal predictive ability. For DeepSurv, Cox-CC, DeepHit, N-MTLR, and Cox-Time, the mean C^{td} -index was 0.773, 0.767, 0.765, 0.732, and 0.659, and the mean IBS was 0.181, 0.192, 0.195, 0.212, and 0.225, respectively. DeepSurv, Cox-CC, and DeepHit yielded comparable performance. Deep learning models are free from the stringent assumptions inherent in standard survival models. Hence, these models are considered flexible alternatives to the standard approaches in scalable, real-world survival problems.

Keywords: *SICU, deep learning, survival analysis, mortality prediction, MIMIC-II*

1. Introduction

The surgical intensive care unit (SICU) is a type of intensive care unit (ICU) for patients undergoing or recovering from surgery. Patients admitted to SICUs are typically in serious health condition, and often have a longer duration of stay, which has been associated with increased mortality rates. Accurately predicting in-hospital mortality and length of stay helps in implementing effective interventions and health care policies, as they are the most significant clinical outcomes for an ICU admission (Hartl et al., 2007; Mosissa et al., 2021).

The Cox proportional hazards (Cox-PH) model has been a salient choice for time-to-event analyses (Wang et al., 2019). However, the underlying linearity and proportional hazards assumptions of this particular model are relatively stringent for scalable, real-world datasets. Deep learning models, free from these assumptions, are emerging as efficient alternatives for the Cox-PH model. These models also have high prediction accuracies that directly assist clinicians in improving treatment plans (Sargent, 2001; Wang et al., 2019; Xiang et al., 2000).

This study applies some of the recently-developed and salient deep learning-based survival models to predict the overall survival of adult patients admitted to SICUs at the Beth Israel Deaconess Medical Center in Boston. The predictive performance of the models was compared using standard performance metrics.

2. Methods

This study involved a secondary data analysis of 2,225 adult patients admitted to SICUs from the MIMIC-II database (Goldberger et al., 2000). The MIMIC database offers the opportunity to develop and validate novel methods for critically ill patients. This database comprises anonymous health-related data of adult patients admitted to critical care units. Every patient is associated with physiological parameters such as cholesterol level, heart rate, serum glucose, O_2 saturation, etc., as well as general descriptors such as gender, age, height, and weight.

The continuous variables were summarized with a mean (SD) or median (Q_1, Q_3) based on the normality of variables, and the categorical variables were summarized with a frequency (%) value.

Authors information:
[a]Department of Data Science, Prasanna School of Public Health, Manipal Academy of Higher Education, Manipal, Karnataka, INDIA. E-mail: lintu.mk@learner.manipal.edu[1]; asha.kamath@manipal.edu[3]
[b]Manipal School of Information Sciences, Manipal Academy of Higher Education, Manipal, Karnataka, INDIA. E-mail: sudarsan.acharya@manipal.edu[2]
[c]Division of Mathematics, School of Advanced Sciences, Vellore Institute of Technology, Chennai, Tamil Nadu, INDIA. E-mail: davidraj.micheal@vit.ac.in[4]
*Corresponding Author: sudarsan.acharya@manipal.edu

Received: November 26, 2021
Accepted: February 16, 2022
Published: October 31, 2022

DOI:https//doi.org/10.22452/mjs.vol41no3.6
Malaysian Journal of Science 41(3): 44-48 (October 2022)

The Kaplan-Meier method was used to visualize the overall survival, and to estimate the mean survival time of patients admitted to SICUs.

For time-to-event analyses of large datasets with several covariates, the standard Cox-PH model often fails, as it was designed for datasets with a relatively smaller number of covariates. Additionally, the Cox-PH model is too simplistic for real-world settings, as it assumes a linear relationship between the outcome and the covariates (Xiang et al., 2000). The proportional hazards assumption is another limitation of the Cox-PH model. According to prior work, it is evident that deep learning models provide more accurate results than standard models such as Cox-PH, without making stringent assumptions. Deep learning models also capture complex interactions between the dependent and independent variables involved, without any distributional assumptions (Sargent, 2001; Wang et al., 2019; Xiang et al., 2000). Hence, deep learning models emerge as efficient alternatives for the Cox-PH model in solving large real-world survival problems.

The five deep learning models employed to predict the survival of SICU patients were: DeepSurv (Katzman et al., 2018), DeepHit (Lee et al., 2018), Cox-CC, Cox-Time (Kvamme et al., 2019), and Neural Multi-Task Logistic Regression (N-MTLR) (Yu et al., 2011). DeepSurv and Cox-CC are non-linear extensions of the Cox model. Cox-Time, DeepHit, and N-MTLR are non-linear and non-proportional extensions of the Cox model. Additionally, DeepHit is able to handle competing risks. Unlike the other models, N-MTLR performs survival analysis through a series of logistic regression models. The neural network-based Cox-PH model based on the linear Cox-PH technique was also fit. All the models were fit using the pycox Python package (Kvamme et al., 2019).

Baseline variables with more than 50% missingness were dropped from the study. Subsequently, the models were built on the 29 baseline variables. The remaining missing observations were imputed using the fancyimpute Python package by employing an iterative imputation method where each feature with missing values is modeled as a function of other features in a round-robin fashion (Rafsunjani et al., 2019).

All of the continuous covariates were standardized, and the categorical variables were binary encoded prior to training the neural network. A five-fold cross-validation was applied with 60% training, 20% validation, and 20% testing sets. For all of the models, we applied the same network architecture used by Kvamme et al (2019), including ReLU activations (Nair & Hinton, 2010), batch normalization (Ioffe & Szegedy, 2015), dropout (Srivastava et al., 2014), and early stopping (Prechelt, 1998). Hyperparameter tuning was performed using a random grid search, and the optimal model was selected based on the scores obtained on the validation set.

The predictive performance of the models was then compared using a time-dependent concordance index (C^{td}-index) (Antolini et al., 2005) and an integrated Brier score (IBS) (Graf et al., 1999). Both measures range between zero and one. The highest C^{td}-index and lowest IBS indicate the best performance. The mean (95% CI) C^{td}-index and IBS were computed from a 5-fold cross-validation procedure.

3. Results

From among the 2,225 patients admitted to SICUs, 331 (14.88%) died in an SICU. The remaining 1,894 (85.12%) patients were considered censored. The median (Q_1, Q_3) duration of stay was 12 (7, 20) days. There were 955 (42.92%) female patients and 1,270 (57.10%) male patients. The mean (SD) age was 60.84 (19.18) years. The baseline characteristics of the two groups (patients who died and survived in SICUs) are described individually in Table 1.

Table 1. Baseline Characteristics.

	Total SICU patients (n = 2225)	Patients died in SICU (n = 331)	Patients survived in SICU (n = 1894)
Age in years (SD)	60.84 (19.18)	71.22 (16.73)	59.03 (19.01)
Height in cm (SD)	171.13 (13.98)	169.40 (11.88)	171.42 (14.02)
Weight in kg (SD)	81.16 (23.02)	80.49 (22.90)	83.25 (23.86)
HR in bpm (SD)	87.72 (19.85)	88.60 (21.85)	89.43 (20.33)
MAP in mmHg (SD)	87.01 (25.90)	86.71 (32.22)	86.14 (23.78)
RR in cpm (SD)	19.26 (6.96)	22.40 (6.88)	19.93 (5.35)
Na in mmol/l (SD)	139.34 (4.11)	139.55 (4.64)	139.31 (4.02)
K in mmol/l (SD)	4.05 (0.65)	4.01 (0.65)	4.06 (0.65)
HCO_3 in mmol/l (SD)	23.25 (4.19)	22.29 (4.59)	23.43 (4.09)
WBC in 103/mm3 (SD)	12.53 (6.09)	12.41 (5.89)	12.41 (5.89)
DiasABP in mmHg (SD)	64.20 (14.92)	62.29 (16.46)	63.84 (14.35)
Glucose in mg/dL(SD)	150.69 (63.42)	163.80 (75.29)	151.22 (59.89)

DOI:https//doi.org/10.22452/mjs.vol41no3.6
Malaysian Journal of Science 41(3): 44-48 (October 2022)

NIDiasABP in mmHg (SD)	62.30 (16.79)	61.02 (17.84)	62.52 (16.60)
NISysABP in mmHg (SD)	127.97 (26.66)	129.63 (32.05)	127.68 (25.62)
NIMAP in mmHg (SD)	82.78 (17.69)	82.70 (19.69)	82.79 (17.33)
Percentage of SaO_2(SD)	96.93 (3.02)	96.20 (1.64)	96 (4.31)
Temperature in °C (SD)	36.65 (1.07)	36.38 (0.63)	36.74 (0.84)
Cholesterol in mg/dL (SD)	160.54 (49.64)	163.38 (58.66)	159.89 (47.59)
FiO_2 (SD)	0.72 (0.25)	0.76 (0.25)	0.71 (0.25)
Mg in mmol/L (SD)	1.79 (0.39)	1.85 (0.37)	1.79 (0.39)
$PaCO_2$ in mmHg (SD)	40.77 (8.99)	38.53 (8.67)	41.24 (8.99)
PaO_2 in mmHg (SD)	188.46 (19.38)	205 (113.35)	184.89 (108.23)
Platelets in cells/nL (SD)	221.01 (100.50)	206.10 (99.75)	223.61 (100.43)
SysABP in mmHg (SD)	132.34 (29.59)	134.63 (35.37)	131.89 (28.30)
pH (SD)	7.37 (0.89)	7.38 (0.10)	7.37 (0.09)
Bilirubin in mg/dl (Q_1,Q_3)	0.80 (0.50, 1.80)	1 (0.50, 2)	0.80 (0.50, 1.70)
Hospital LOS in days (Q_1,Q_3)	12 (7, 20)	9 (5, 17)	13 (8, 21)
BUN in mg/dL(Q_1,Q_3)	17 (13, 27)	22 (14, 37)	16 (12, 25)
Creatinine in mg/dL(Q_1,Q_3)	0.90 (0.70, 1.30)	1.10 (0.80, 1.70)	0.90 (0.70, 1.20)
Lactate in mmol/L (Q_1,Q_3)	2.20 (1.40, 3.70)	2.80 (1.70, 4.20)	2 (3.60, 1.30)
Gender – male (%)	1270 (57.10)	183 (14.41)	1087 (85.59)

Continuous variables are presented as mean (SD), median (Q_1, Q_3); categorical variables are presented as frequency (%)

The survival probability was visualized as a function of time using the Kaplan-Meier method, as shown in Figure 1. The mean (95% CI) survival time of the patients admitted to SICUs was 95.09 (83.36, 106.83) days.

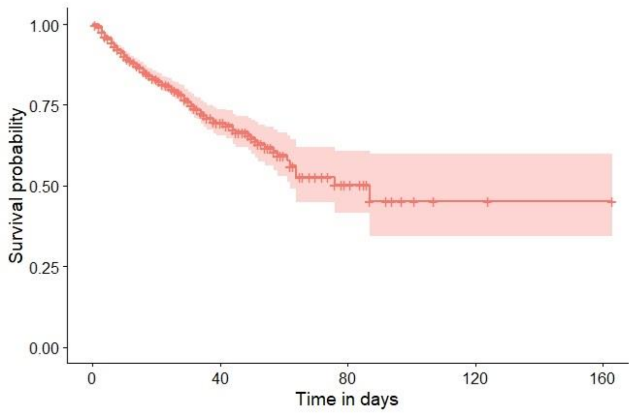

Figure 1. Survival probability of patients admitted to SICU.

The mean (95% CI) of the Ctd-index and IBS from a 5-fold cross-validation procedure for the different models are listed in Table 2.

Table 2. Performance of Deep Learning Models in Terms of C^{td}-index and IBS.

Model	C^{td}-index (95% CI)	IBS (95% CI)
Cox-CC	0.767 (0.737,0.798)	0.192 (0.106,0.279)
Cox-Time	0.659 (0.611,0.707)	0.225 (0.173,0.278)
DeepSurv	0.773 (0.743,0.803)	0.181 (0.139,0.222)
DeepHit	0.765 (0.727,0.804)	0.195 (0.181,0.208)
N-MTLR	0.732 (0.706,0.757)	0.212 (0.174,0.251)
Cox-PH	0.673 (0.631,0.717)	0.220 (0.173,0.277)

Among the five models, DeepSurv achieved the most optimal predictive performance based on both performance metrics; its mean C^{td}-index was 0.773 (95% CI: 0.743-0.803), and its mean integrated Brier score was 0.181 (95% CI: 0.139-0.222). Compared to all of the other models, Cox-Time obtained the least performance (Ctd-index 0.659, 95% CI 0.611, 0.707; Brier score 0.225, 95% CI 0.173, 0.278). The performance of both Cox-CC and DeepHit was similar to that of DeepSurv. For all of the models, both performance metrics resulted in similar results.

DOI:https//doi.org/10.22452/mjs.vol41no3.6
Malaysian Journal of Science 41(3): 44-48 (October 2022)

4. Discussion

This study utilized the data extracted from the MIMIC-II database to evaluate the performance of various deep learning-based models to predict the survival of patients admitted to SICUs. This involved identifying a suitable algorithm to bring awareness about potential alternatives to the traditional Cox-PH model. Using such promising alternatives, clinicians could make use of the readily-available large-scale and complex real-world datasets to enforce enhanced treatment policies (Sargent, 2001; Wang et al., 2019; Xiang et al., 2000). The improved predictive ability of such models has already been proven in the existing literature (Katzman et al., 2018; Kvamme et al., 2019; Lee et al., 2018; Yu et al., 2011), in the context of large-scale healthcare record databases.

The results suggest that the DeepSurv model predicts the survival of SICU patients better than the other models. As predicting in-hospital mortality rates and quantifying patient health are vital in critical care research, using an appropriate model for a given dataset has a great impact (Austin et al., 2002; Jalali et al., 2020; Yun et al., 2021).

The limitations of his study include the use of secondary data and the large proportion of missing observations. To the best of our knowledge, this is the first study that attempts to evaluate the performance of various deep learning-based survival models to predict the survival of critically-ill patients admitted to SICUs. The large sample size and number of predictors in the SICU dataset from the MIMIC-II database further strengthens the conclusions of our study.

5. Conclusions

The Standard Cox-PH model underperforms for large datasets, attributable to its restrictive model assumptions. Deep learning-based models that are free from such assumptions are excellent alternatives for the survival analysis of large real-world problems. Deep learning-based models are typically computationally-expensive, but with suitable tuning, their superior predictive performance can be leveraged for the accurate prediction of in-hospital mortality rates and in turn the effective management of ICU patients.

6. Acknowledgements

We thank the team behind the MIMIC database for providing all the data used in this study.

7. References

Antolini, L., Boracchi, P., & Biganzoli, E. (2005). A time-dependent discrimination index for survival data. Statistics in medicine 24: 3927-3944.

Austin, P. C., Rothwell, D. M., & Tu, J. V. (2002). A Comparison of Statistical Modeling Strategies for Analyzing Length of Stay after CABG Surgery. Health Services and Outcomes Research Methodology 3: 107-133.

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online] 101, pp. e215–e220.

Graf, E., Schmoor, C., Sauerbrei, W., & Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. Statistics in medicine 18: 2529-2545.

Hartl, W. H., Wolf, H., Schneider, C. P., Küchenhoff, H., & Jauch, K. W. (2007). Acute and long-term survival in chronically critically ill surgical patients: a retrospective observational study. Critical Care 11: 1-11.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International conference on machine learning, pp. 448-456.

Jalali, A., Lonsdale, H., Do, N., Peck, J., Gupta, M., Kutty, S., ... & Ahumada, L. M. (2020). Deep Learning for Improved Risk Prediction in Surgical Outcomes. Scientific Reports 10: 9289-9302.

Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., & Kluger, Y. (2018). DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC medical research methodology 18: 24-36.

Kvamme, H., Borgan, Ø., & Scheel, I. (2019). Time-to-event prediction with neural networks and Cox regression. Journal of Machine Learning Research 20: 1-30.

Lee, C., Zame, W. R., Yoon, J., & van der Schaar, M. (2018). Deephit: A deep learning approach to survival analysis with competing risks. In: Thirty-second AAAI conference on artificial intelligence.

DOI:https//doi.org/10.22452/mjs.vol41no3.6
Malaysian Journal of Science 41(3): 44-48 (October 2022)

Mosissa, D., Alemu, S., Rad, M. H., & Yesuf, E. A. (2021). Outcomes of Surgical Patients Admitted to the Intensive Care Unit of Jimma University Medical Center. Health Science Journal 15: 1-4.

Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In: 27th International Conference on Machine Learning: USA. Omnipress, pp. 807-814.

Prechelt, L. (1998). Early stopping-but when?. In: Neural Networks: Tricks of the trade, pp. 55-69. Springer, Berlin, Heidelberg.

Rafsunjani, S., Safa, R. S., Al Imran, A., Rahim, M. S., & Nandi, D. (2019). An empirical comparison of missing value imputation techniques on APS failure prediction. International Journal of Information Technology and Computer Science 2: 21-29.

Sargent, D. J. (2001). Comparison of artificial neural networks with other statistical approaches: results from medical data sets. Cancer: Interdisciplinary International Journal of the American Cancer Society 91: 1636-1642.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research 15: 1929-1958.

Wang, P., Li, Y., & Reddy, C. K. (2019). Machine learning for survival analysis: A survey. ACM Computing Surveys 51: 1-36.

Xiang, A., Lapuerta, P., Ryutov, A., Buckley, J., & Azen, S. (2000). Comparison of the performance of neural network methods and Cox regression for censored survival data. Computational statistics & data analysis 34: 243-257.

Yu, C. N., Greiner, R., Lin, H. C., & Baracos, V. (2011). Learning patient-specific cancer survival distributions as a sequence of dependent regressors. Advances in neural information processing systems 24: 1845–1853.

Yun, K., Oh, J., Hong, T. H., & Kim, E. Y. (2021). Prediction of Mortality in Surgical Intensive Care Unit Patients Using Machine Learning Algorithms. Frontiers in Medicine 8: 406-415

DOI:https//doi.org/10.22452/mjs.vol41no3.6
Malaysian Journal of Science 41(3): 44-48 (October 2022)