



Research Article

Application of the DeepSurv Model to Predict Survival in Patients with Kidney Failure Undergoing Hemodialysis

Rizki Amanda¹; Aviolla Terza Damaliana²; Mohammad Idhom³

¹ Universitas Pembangunan Nasional “Veteran” Jawa Timur, Surabaya, Indonesia, 22083010045@student.upnjatim.ac.id

² Universitas Pembangunan Nasional “Veteran” Jawa Timur, Surabaya, Indonesia, aviolla.terza.sada@upnjatim.ac.id

³ Universitas Pembangunan Nasional “Veteran” Jawa Timur, Surabaya, Indonesia, idhom@upnjatim.ac.id

Correspondence should be addressed to Rizki Amanda; 22083010045@student.upnjatim.ac.id

Received 03 Jan 2026; Accepted 15 Februari 2026; Published 30 March 2026

© Authors 2026. CC BY-NC 4.0 (non-commercial use with attribution, indicate changes).

License: <https://creativecommons.org/licenses/by-nc/4.0/> — Published by Indonesian Journal of Data and Science.

Abstract:

This study aims to improve survival prediction in patients with kidney failure undergoing hemodialysis, given their high mortality risk. Traditional models such as Cox Proportional Hazards (Cox PH) have limitations in capturing complex and nonlinear relationships in clinical data. Therefore, this study applies DeepSurv, a deep learning-based survival model, and compares its performance with Cox PH and Cox PH Spline. A total of 300 patients were included, with 165 events and 135 censored observations. The data were split into training and testing sets. DeepSurv was implemented using two hidden layers (64 and 32 neurons), a dropout rate of 0.2, and a learning rate of $1e-3$. The model was trained for up to 1000 epochs with early stopping at epoch 435. Performance was evaluated using the concordance index (C-index) and time-dependent AUC at 365, 544, and 730 days. Patients were stratified into low-, medium-, and high-risk groups based on predicted scores. Results showed that Cox PH achieved a C-index of 0.913 and average AUC of 0.964, while Cox PH Spline reached 0.917 and 0.971. DeepSurv achieved a C-index of 0.920 and average AUC of 0.969. Performance differences were small, but DeepSurv provided consistent individual risk estimates. In conclusion, DeepSurv is a flexible approach with performance comparable to Cox-based models. Further external validation and clinical evaluation are needed before wider application.

Keywords: DeepSurv; Hemodialysis; Kidney Failure; Survival Analysis.

Dataset link: <https://drive.google.com/file/d/17WxI5WNKR8L2chqQJBjLdjry61VIR9tZ/view?usp=sharing>

1. Introduction

Chronic Kidney Disease (CKD) is a major global health problem with a prevalence that continues to increase each year. CKD is defined as a progressive and irreversible decline in kidney function, characterized by a glomerular filtration rate (GFR) below 60 ml/min for more than three months or persistent structural kidney damage [1]. In advanced stages, CKD progresses to end stage renal disease (ESRD), which requires renal replacement therapy such as hemodialysis, peritoneal dialysis, or kidney transplantation to sustain life [2]. According to the National Kidney Foundation, the average life expectancy of patients undergoing hemodialysis is only about 5-10 years despite optimal care [3].

Globally, the World Health Organization (WHO) reports approximately 850,000 deaths annually due to CKD, making it one of the leading causes of death worldwide, while the Global Burden of Disease Study (2019) estimates more than 1.4 million deaths related to CKD. In Indonesia, the prevalence of CKD has also increased significantly, with the 2018 Basic Health Research (Riskesdas) reporting more than 739,000 cases [4]. The growing number of hemodialysis patients poses a major challenge to healthcare systems, particularly in developing countries with limited resources. In Indonesia, dialysis cases continue to rise each year, placing a heavy burden on hospitals and the national health system.

Patients undergoing hemodialysis face a high risk of mortality due to various complications and comorbidities. Factors such as age, sex, dialysis frequency, hypertension, hypotension, and diabetes mellitus contribute to the complexity of survival outcomes. The presence of heterogeneous patient characteristics and dynamic clinical conditions further complicates survival prediction. Therefore, early identification of high-risk patients is crucial to improve clinical management and reduce mortality rates.

Conventional survival analysis methods, such as Kaplan Meier estimation and Cox Proportional Hazards (Cox PH) regression, are widely used to estimate survival probabilities and identify prognostic factors [5]. However, these approaches rely on assumptions of linear covariate effects and proportional hazards, which often do not hold in heterogeneous and complex hemodialysis clinical data [6], [7], [8]. As a result, traditional models may fail to capture non linear relationships and complex interactions among clinical variables, leading to less accurate individual survival predictions [9]. This limitation highlights a clear research gap in developing more flexible models capable of handling such complexities.

Recent advances in machine learning and deep learning have introduced more flexible approaches for survival prediction [10]. DeepSurv extends the Cox PH model by utilizing deep neural networks to learn nonlinear relationships between covariates and mortality risk, thereby enabling more personalized risk estimation [11]. Previous studies have demonstrated that DeepSurv can outperform conventional Cox based models in predictive accuracy and risk stratification, particularly in complex clinical datasets [12], [13].

Based on this gap, this study hypothesizes that DeepSurv can improve long term discrimination performance compared to conventional Cox PH and Cox PH Spline models in predicting survival outcomes of hemodialysis patients [14], [15]. To evaluate this hypothesis, this study applies DeepSurv to generate individual risk scores and stratifies patients into low, medium, and high-risk groups. The performance of DeepSurv is systematically compared with Cox PH and Cox PH Spline models to assess predictive capability and the quality of risk stratification [16], [17].

This study uses hemodialysis patient data from RSUD Haji Surabaya and aims to contribute to the development of more flexible, interpretable, and clinically applicable survival prediction models. The findings are expected to support clinical decision making and provide insights into the application of deep learning-based survival models in healthcare settings, particularly in the context of risk assessment.

2. Method

This study proposes a systematic survival analysis framework to identify mortality risk levels in patients with chronic kidney disease undergoing hemodialysis at RSUD Haji Surabaya. The workflow includes clinical data preprocessing, survival modeling using both statistical and deep learning approaches, and comprehensive model evaluation.

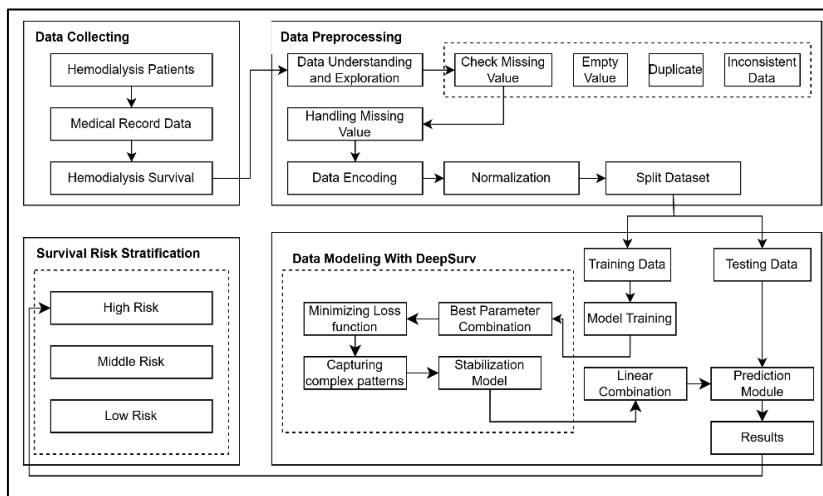


Figure 1. Research Flowchart

Retrospective medical records of 300 patients with chronic kidney disease undergoing hemodialysis were collected from RSUD Haji Surabaya, covering the period from January 1, 2022, to December 1, 2025. Independent variables consisted of demographic and clinical factors such as age, sex, dialysis frequency, hypertension, diabetes, hemoglobin levels, and other relevant medical indicators. The dependent variable was survival time, defined as the duration from the initiation of hemodialysis to the occurrence of death. Patients who did not experience the event during the observation period were treated as right censored cases [18], [19]. Missing values were handled through appropriate preprocessing steps, including imputation for numerical variables and consistency checks for categorical variables.

Table 1. Kidney Failure Dataset

No	Patient Status	Age	Dialysis Frequency	...	Diabetes	Urea	Time to Event (Days)
1	1	47	4	...	0	17	386
2	0	55	8	...	1	77	1411
3	0	41	8	...	1	47	1411
...
98	0	79	8	...	1	23	352
...
300	1	58	1	...	1	42	60

The dataset was divided using a holdout strategy into 80% training data (240 patients) and 20% testing data (60 patients). The split was performed randomly with a fixed random seed to ensure reproducibility, while maintaining the proportion of event and censored observations.

Survival outcomes are represented by survival time T_i and event status E_i , while clinical predictors $(X_1, X_2, \dots, X_{10})$ form the covariate matrix used in the DeepSurv model to estimate individual mortality risk.

The primary predictive approach in this study is DeepSurv, a deep feed forward neural network that extends the Cox Proportional Hazards (CPH) model by learning complex nonlinear interactions among clinical variables [20], [21]. Unlike the classical Cox model, which assumes linear covariate effects, DeepSurv replaces the linear risk function with a neural network representation. The hazard function is expressed as:

$$h(t | \mathbf{X}) = h_0(t) \cdot \exp^{f(x_i; \theta)} \quad (1)$$

where $f(x_i; \theta)$ denotes the network output (log risk score) learned from patient covariates such as age, blood pressure, creatinine levels, comorbidities, and dialysis frequency. The DeepSurv architecture is constructed using multiple hidden layers with the Scaled Exponential Linear Unit (SELU) activation function, chosen for its ability to maintain activation stability and accelerate convergence. To reduce the risk of overfitting, regularization techniques such as dropout and weight decay are applied. Model training is performed by optimizing network parameters through hyperparameter tuning. In this study, the network consists of two hidden layers with 64 and 32 neurons, respectively, a dropout rate of 0.2, and is trained using the Adam optimizer with a learning rate of 1×10^{-3} . The model is trained for a maximum of 1000 epochs with early stopping based on validation loss to prevent overfitting. The loss function is defined as the negative log partial likelihood from the Cox framework, thereby preserving the fundamental structure of survival analysis while adopting a deep learning approach [22], [23].

For comparison, the DeepSurv approach is evaluated against the Cox Proportional Hazards and Cox PH Spline models. The Cox PH model is used as a conventional survival analysis benchmark. The Cox PH Spline model is applied to accommodate nonlinear relationships in specific variables [24], [25]. Based on exploratory analysis, dialysis frequency showed indications of a complex nonlinear relationship with survival outcomes; therefore, a spline transformation with 3 degrees of freedom ($df = 3$) is applied to this variable. Knot placement is determined using a quantile-based approach to ensure stable estimation.

Model performance is primarily evaluated using the Concordance Index (C-index), which assesses the model's ability to correctly rank patients according to survival risk while accounting for censored observations [26], [27]. The C-index is defined as:

$$C - Index = \frac{N_{konkordan} + 0.5 \times N_{tie}}{N_{evaluasi}} \quad (2)$$

where $N_{konkordan}$ is the number of correctly ordered pairs, N_{tie} is the number of pairs with identical predicted risk, and $N_{evaluasi}$ is the total number of comparable pairs. A higher C-index indicates better discriminative performance of the survival model.

In addition, time dependent ROC curves and AUC are used to measure model discrimination at specific time horizons. To complement discrimination performance, model calibration is also evaluated using calibration curves and the Integrated Brier Score (IBS), ensuring that predicted survival probabilities are consistent with observed outcomes [28], [29].

Individual risk scores generated by the DeepSurv model are subsequently used for clinical risk stratification. Based on the distribution of risk scores, patients are categorized into three groups: low risk, medium risk, and high risk. To evaluate differences in survival among these risk groups, Kaplan Meier survival curves are estimated for each group. Divergence among the survival curves serves as an indicator of the effectiveness of the model's risk stratification [30], [31].

Thus, this study not only assesses survival prediction performance in terms of discrimination but also emphasizes calibration and the model's effectiveness in practically and interpretably classifying patient mortality risk levels.

3. Result and Discussion

The baseline characteristics of patients with chronic kidney disease undergoing hemodialysis are summarized in **Table 2**. The table presents demographic and clinical variables for the overall cohort as well as for the training and test subsets. Continuous variables are reported as medians with interquartile ranges (Q1–Q3), while categorical variables are presented as frequencies and percentages, with p-values indicating differences between groups.

Table 2. Baseline Characteristics of Hemodialysis Patients

Variable	Overall	Training	Test	p-value
Age (median [Q1-Q3])	58.0 [51.0–66.0]	58.5 [51.0–66.2]	55.5 [47.8–64.0]	0.332
Hemodialysis Frequency (median [Q1–Q3])	5.0 [1.0–7.0]	5.0 [1.0–7.0]	5.0 [1.8–7.0]	0.896
Hemoglobin (g/dL) (median [Q1–Q3])	9.3 [8.1–10.9]	9.3 [8.1–10.9]	9.1 [8.0–10.9]	0.550
Creatinine (mg/dL) (median [Q1–Q3])	4.7 [2.8–7.7]	4.5 [2.8–7.7]	5.0 [2.7–7.8]	0.690
Urea (mg / dl) (median [Q1–Q3])	45.0 [28.8–69.5]	45.0 [28.0–67.0]	49.5 [31.5–73.8]	0.359
Systolic Blood Pressure (median [Q1–Q3])	140.5 [121.0–165.0]	142.0 [125.0–165.0]	137.0 [110.8–174.8]	0.235
Sex (n, %)				0.665
0	150 (50.0%)	122 (50.8%)	28 (46.7%)	
1	150 (50.0%)	118 (49.2%)	32 (53.3%)	
Diabetes (n, %)				0.348
0	44 (14.7%)	38 (15.8%)	6 (10.0%)	
1	256 (85.3%)	202 (84.2%)	54 (90.0%)	
Hypertension (n, %)				0.061

Variable	Overall	Training	Test	p-value
0	150 (50.0%)	113 (47.1%)	37 (61.7%)	
1	150 (50.0%)	127 (52.9%)	23 (38.3%)	
Hypotension (n, %)				0.818
0	289 (96.3%)	232 (96.7%)	57 (95.0%)	
1	11 (3.7%)	8 (3.3%)	3 (5.0%)	

Baseline characteristics were generally comparable between the training and test sets, with no statistically significant differences observed for most variables ($p > 0.05$). Clinical measures such as age, laboratory parameters, and hemodialysis frequency exhibited consistent distributions, indicating a well-balanced cohort split. These results suggest that subsequent survival modeling can be conducted without significant baseline bias between the datasets.

For the Cox Proportional Hazards model, statistical inference was further supported using the Wald test to assess individual covariate significance, the proportional hazards assumption test to verify model validity, and the log likelihood test to evaluate overall model fit.

Subsequently, a feature importance analysis was performed to evaluate the relative contribution of each predictor in the standard Cox Proportional Hazards model, the Cox model with spline transformation, and the DeepSurv model, in order to assess the consistency of survival patterns across datasets.

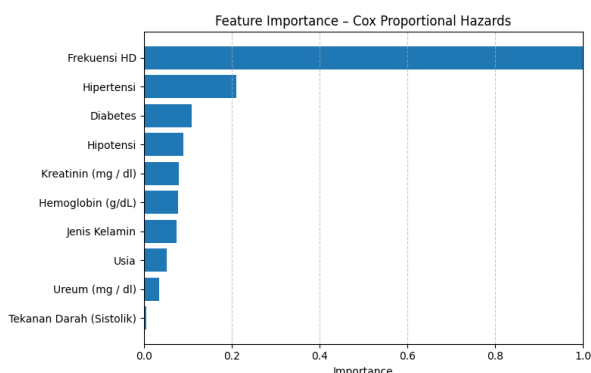


Figure 2. Feature Importance – Cox Proportional Hazard

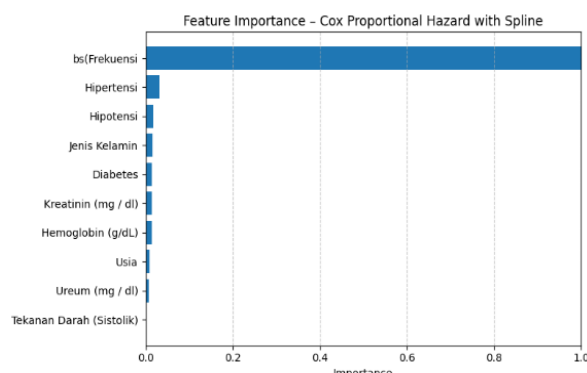


Figure 3. Feature Importance – Cox Proportional Hazard Spline

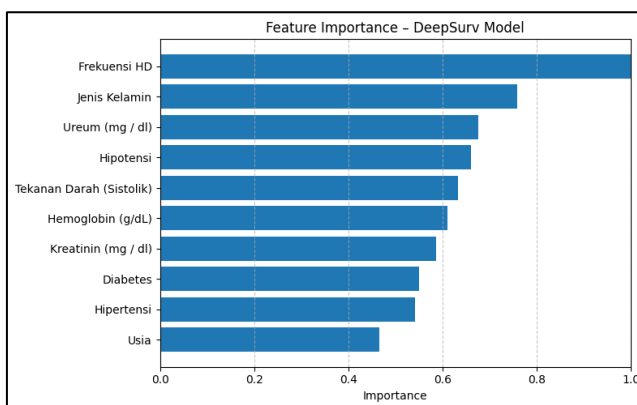


Figure 4. Feature Importance – DeepSurv Model

In the standard Cox Proportional Hazards model (Figure 2), hemodialysis frequency emerged as the most influential predictor of survival, showing a markedly stronger effect than the other covariates, while hypertension, diabetes, and hypotension exhibited moderate contributions. When spline transformations were applied (Figure 3), the nonlinear representation of hemodialysis frequency remained the dominant feature, whereas the relative importance of other clinical variables further decreased.

In contrast, the DeepSurv model (Figure 4) demonstrated a more distributed pattern of feature importance, in which multiple variables including hemodialysis frequency, sex, urea levels, blood pressure, and other laboratory parameters contributed meaningfully to survival prediction.

Feature importance in the DeepSurv model was derived from the magnitude of the neural network weights in the first hidden layer. Specifically, the importance of each feature was computed as the mean absolute value of its corresponding input weights, followed by normalization. From a theoretical perspective, this approach reflects how strongly each input variable influences the learned risk representation, where larger weight magnitudes indicate greater contribution to the predicted log risk function. Although this method provides an interpretable approximation, it should be noted that it does not fully capture complex nonlinear interactions as in permutation-based methods.

Kaplan Meier curves were used to compare survival distributions between the training and test datasets for the standard Cox Proportional Hazards, Cox spline, and DeepSurv models.

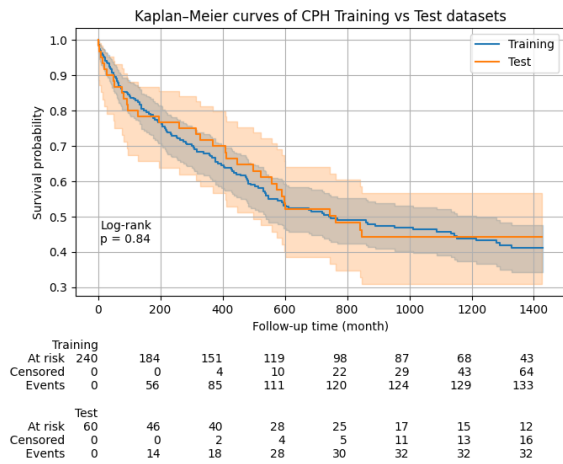


Figure 5. Kaplan Meier Curves – Cox Proportional Hazard

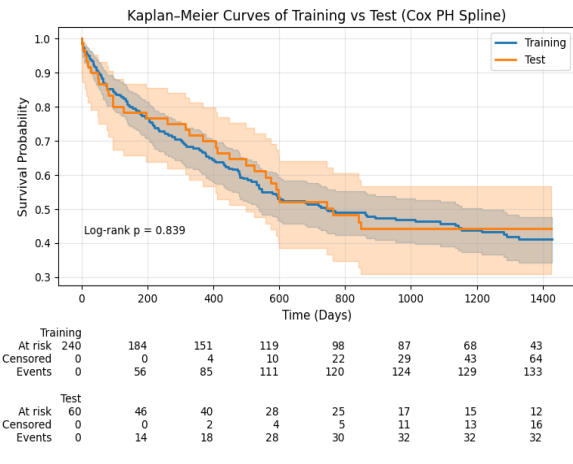


Figure 6. Kaplan Meier Curves – Cox Proportional Hazard Spline

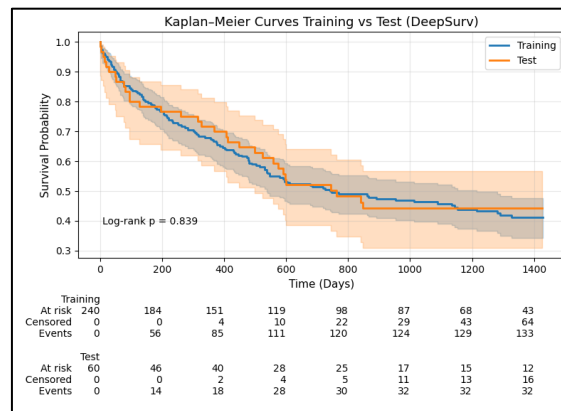


Figure 7. Kaplan Meier Curves – DeepSurv Model

For the standard Cox model, the survival curves of the training set (n = 240) and test set (n = 60) largely overlapped, with no significant differences observed (log rank p = 0.84), indicating comparable baseline survival patterns. Similar results were obtained for the Cox spline model, where the inclusion of non linear effects did not alter the alignment of survival curves between datasets (log rank p = 0.839).

For the DeepSurv model, Kaplan Meier analysis likewise showed no statistically significant differences between the training and test datasets (log rank p = 0.839). In the training set, the cumulative number of events increased from 56 at day 200 to 133 at day 1400, while the number of patients at risk declined from 240 at baseline to 43 at the end of the observation period. A similar pattern was observed in the test set, with cumulative events rising from 14 at day 200 to 32 at day 1400, and 12 patients remaining at risk.

Overall, the consistent overlap of Kaplan Meier curves and non-significant log rank test results across all models indicate good dataset comparability and support the robustness of subsequent survival modeling.

Following the comparison of survival distributions between the training and test datasets, further analysis was conducted using Kaplan Meier curves stratified by predicted risk groups (low, medium, and high risk) to evaluate the discriminative ability of each survival model. Risk stratification was performed separately for the standard Cox Proportional Hazards model, the Cox model with spline transformation, and the DeepSurv model, with the corresponding Kaplan Meier curves presented in **Figures 8-10**.

Risk groups were defined based on tertile cutoffs of the predicted risk scores, where patients in the lowest 33% were categorized as low risk, the middle 33% as medium risk, and the highest 33% as high risk. This approach ensures balanced group sizes and is commonly used in survival analysis to facilitate clinical interpretability.

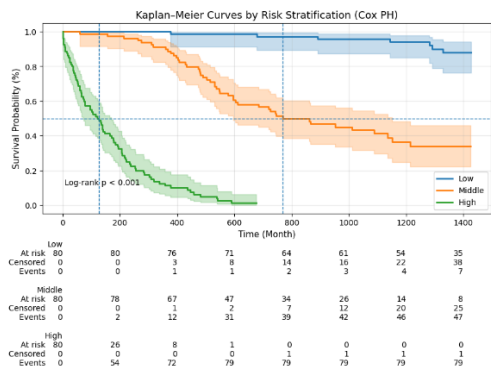


Figure 8. Risk Stratification– Cox Proportional Hazard

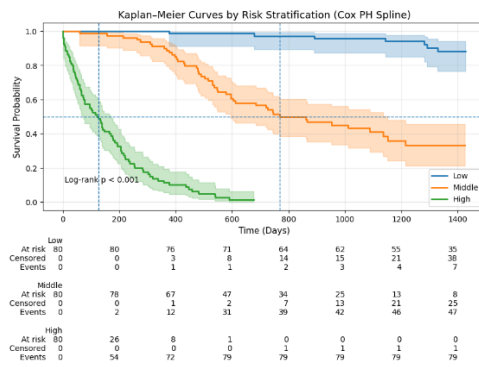


Figure 9. Risk Stratification – Cox Proportional Hazard Spline

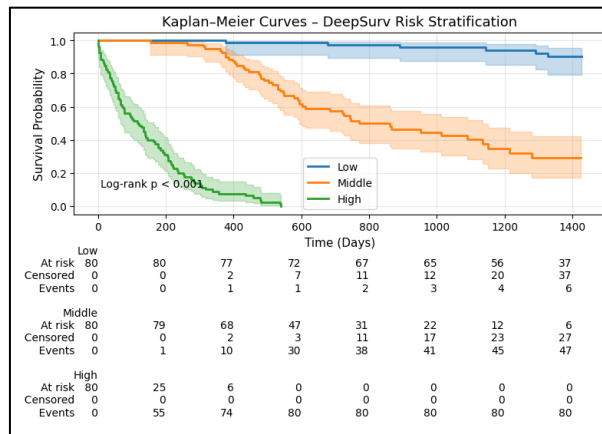


Figure 10. Risk Stratification – DeepSurv Model

The Kaplan Meier curves for the Cox Proportional Hazards model demonstrate clear separation among the three risk groups, with a statistically significant log rank test ($p < 0.001$). The low risk group ($n = 80$) recorded 7 events at day 1400 with 35 patients remaining at risk, the medium risk group recorded 47 events, while the high risk group experienced early events, with 54 events by day 200 and 79 events by day 600. The Cox Proportional Hazards model with spline transformation exhibited a similar stratification pattern ($p < 0.001$), in which the low risk group still recorded 7 events at day 1400, the medium risk group 47 events, and the high risk group 79 events by day 600, indicating that the inclusion of spline terms preserved the model's discriminative ability.

In the DeepSurv model, the Kaplan Meier curves also showed clear risk separation ($p < 0.001$). The low risk group recorded 6 events at day 1400 with 37 patients remaining at risk, the medium risk group 47 events, and the high risk group exhibited the most aggressive risk profile, with 55 events by day 200 and all events (80 events) occurring by day 600.

After individual risk scores were obtained from the DeepSurv model, survival probabilities were estimated at time horizons of 365, 544, and 730 days. Based on the risk score values, patients were categorized into low, medium, and high risk groups. The following table presents the risk scores, risk group assignments, and estimated survival probabilities at each time horizon.

Table 3. Probability of Survival

	Risk Score	Risk Group	Probability of Survival		
			365 Days	544 Days	730 Days
0	3.3490	High	10.87%	0.03%	0%
1	-3.3081	Middle	99.71%	98.99%	97.60%
2	-0.9077	Middle	96.90%	89.49%	76.51%
3	-5.6527	Low	99.97%	99.90%	99.76%
4	-2.7680	Middle	99.51%	98.28%	95.92%
...

Table 3 shows that patients in the high risk group have very low survival probabilities, with $S_{365} = 0.109$ decreasing to nearly zero by days 544 and 730. In contrast, the low risk group maintains a very high survival probability up to day 730 ($S_{730} = 0.998$). The medium risk group exhibits intermediate survival values, with a gradual decline from $S_{365} = 0.969$ to $S_{730} = 0.765$, confirming the effectiveness of the DeepSurv model in patient risk stratification.

The predictive accuracy of the model at specific time horizons was evaluated using time dependent area under the curve (AUC) at days 365, 544, and 730 to assess short to long term risk classification performance.

Table 4. Time Dependent

Model	Time Dependent AUC		
	365 Days	544 Days	730 Days
Cox Proportional Hazard	0.958	0.972	0.973
Cox Proportional Hazard With Spline	0.968	0.979	0.969
DeepSurv	0.964	0.974	0.976

Predictive performance at specific time horizons was evaluated using time dependent AUC at days 365, 544, and 730. The Cox Proportional Hazards model with spline transformation achieved the highest AUC values at day 365 (0.968) and day 544 (0.979), while the DeepSurv model produced consistently high AUC values across all evaluation horizons, with the highest value at day 730 (0.976).

However, it is important to note that the differences in AUC values across models are relatively small. Therefore, strong claims of superiority should be interpreted with caution, and future work should incorporate statistical uncertainty estimation, such as confidence intervals or bootstrap resampling, to provide more rigorous model comparisons.

To complement the evaluation of model discrimination, calibration performance was assessed using the Integrated Brier Score (IBS). The IBS measures the average squared difference between predicted survival probabilities and observed outcomes over time, while accounting for censored data. Lower IBS values indicate better agreement between predicted and actual survival, reflecting more accurate probability estimation. The IBS was computed over a restricted time grid within the observed follow up range of the test dataset to ensure valid and comparable evaluation across models.

Table 5. Integrated Brier Score

Model	Integrated Brier Score
Cox Proportional Hazard	0.10796
Cox Proportional Hazard With Spline	0.07034
Deepsurv	0.06816

Based on the results presented in [Table 5](#), the DeepSurv model achieved the lowest IBS value (0.06816), followed closely by the Cox Proportional Hazards model with spline transformation (0.07034), while the standard Cox Proportional Hazards model showed the highest IBS value (0.10796).

These findings indicate that both DeepSurv and the spline-based Cox model provide better calibration performance compared to the standard Cox model, with only a small difference between the two best performing approaches. The improvement observed in the spline based Cox model suggests that incorporating nonlinear effects enhances the model's ability to estimate survival probabilities more accurately.

Meanwhile, the DeepSurv model demonstrates the best overall calibration performance, indicating its effectiveness in capturing complex relationships among covariates and translating them into more accurate survival probability estimates. However, the differences between DeepSurv and the Cox spline model are relatively small, suggesting that both approaches perform comparably in terms of calibration. Therefore, strong claims of superiority should be interpreted with caution, and further statistical validation, such as confidence intervals or bootstrap analysis, may be required to confirm the significance of these differences.

The discriminative performance of the models was evaluated using the concordance index (C-index) to assess their ability to rank patients according to survival risk while accounting for censored data.

Table 6. Model Evaluation

Model	C-Index	
	Training Dataset	Testing Dataset
Cox Proportional Hazard	0.911	0.913
Cox Proportional Hazard With Spline	0.911	0.917
DeepSurv	0.925	0.920

Model performance evaluation using the concordance index (C-index) indicates that all models exhibit high discriminative ability, with C-index values exceeding 0.90 on both training and test datasets. The standard Cox Proportional Hazards model achieved a C-index of 0.911 on the training data and 0.913 on the test data, indicating stable and consistent performance. The Cox Proportional Hazards model with spline transformation yielded the same

C-index on the training data (0.911) and an improved value on the test data (0.917), reflecting slightly better model generalization. Among all approaches, the DeepSurv model attained the highest C-index on the training data (0.925) and maintained strong performance on the test data (0.920), demonstrating the overall best risk discrimination capability.

Following risk stratification, survival curve comparisons were conducted to further evaluate the predictive performance of each model. **Figure 11** presents the survival curves generated by the standard Cox PH, Cox PH spline, and DeepSurv models, with Kaplan Meier estimates used as the baseline.

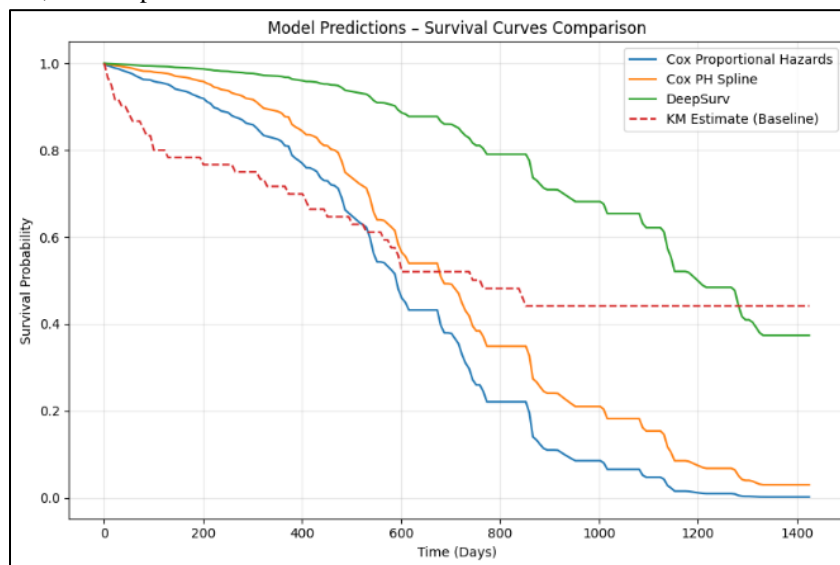


Figure 11. Model Prediction – Survival Curves Comparison

The visualization shows that the DeepSurv curve maintains slightly higher survival probabilities compared with the standard Cox PH and Cox PH spline models, suggesting its ability to capture more flexible survival patterns. The standard Cox PH curve exhibits a more rapid decline in survival, while the Cox PH spline model provides moderate improvement. Kaplan Meier estimates serve as an empirical reference, indicating that all models generally follow the observed data trends.

Overall, while DeepSurv demonstrates consistent and stable performance, the differences compared to Cox based models remain relatively small, highlighting the importance of balancing model complexity with interpretability in clinical applications.

4. Conclusion

Based on the analysis results, all three models the standard Cox Proportional Hazards, Cox PH spline, and DeepSurv were able to classify hemodialysis patients into low, medium, and high risk groups. The Cox based models demonstrated clear risk separation, with hemodialysis frequency emerging as the primary predictor, whereas DeepSurv captured more diverse and non linear contributions from multiple clinical variables. Kaplan Meier analysis confirmed good agreement between the training and test datasets (log rank $p > 0.8$), indicating consistent survival patterns across data splits.

In terms of predictive performance, all models showed strong discrimination ability. The DeepSurv model achieved slightly higher C-index values (0.925 for training and 0.920 for testing) compared to the Cox based models, along with consistently high time dependent AUC values across the evaluated time horizons (365, 544, and 730 days).

However, the differences in performance metrics between models were relatively small, suggesting that all approaches provide comparable discrimination capability.

Overall, DeepSurv demonstrates the ability to model complex relationships within clinical data and supports effective risk stratification. Nevertheless, these findings should be interpreted cautiously. This study is based on retrospective data from a single center, which may introduce location specific bias and limit generalizability. In addition, no external validation was performed. Therefore, further studies using multicenter datasets and external validation are necessary before considering clinical implementation or broader generalization of the proposed model.

References:

- [1] M. Muhajir and C. H. Imanina, "Survival analysis of chronic kidney disease patients using stratified cox regression in arifin achmad hospital, Pekanbaru, Riau," *Indian J. Public Health Res. Dev.*, vol. 9, no. 11, p. 1592, 2018, doi: [10.5958/0976-5506.2018.01673.X](https://doi.org/10.5958/0976-5506.2018.01673.X).
- [2] A. Séraphin *et al.*, "Survival and Predictors of Mortality in Hemodialysis Patients at the Departmental Teaching Hospital of Borgou and Alibori (Benin)," *Int. J. Nephrol. Kidney Fail.*, vol. 11, no. 2, 2025, doi: [10.16966/2380-5498.254](https://doi.org/10.16966/2380-5498.254).
- [3] T. Tourountzis *et al.*, "Survival and Predictors of Mortality in Hemodialysis Patients," *Nephrology Dialysis Transplantation*, vol. 38, no. Supplement_1, Jun. 2023, doi: [10.1093/ndt/gfad063c_4742](https://doi.org/10.1093/ndt/gfad063c_4742).
- [4] R. Laksana Aryananda, Trimono, W. Syaifullah J, and W. S. Wan Awang, "Hyperparameter optimization of XGBoost using artificial bee colony for predicting medical complications in hemodialysis patients," *Jurnal Ilmiah Kursor*, vol. 13, no. 1, pp. 32–45, Jul. 2025, doi: [10.21107/kursor.v13i1.459](https://doi.org/10.21107/kursor.v13i1.459).
- [5] Y. Huang, J. Li, M. Li, and R. R. Aparasu, "Application of machine learning in predicting survival outcomes involving real-world data: a scoping review," *BMC Med. Res. Methodol.*, vol. 23, no. 1, p. 268, Nov. 2023, doi: [10.1186/s12874-023-02078-1](https://doi.org/10.1186/s12874-023-02078-1).
- [6] J. Lei *et al.*, "The predictive value of modified-DeepSurv in overall survivals of patients with lung cancer," *iScience*, vol. 26, no. 11, p. 108200, Nov. 2023, doi: [10.1016/j.isci.2023.108200](https://doi.org/10.1016/j.isci.2023.108200).
- [7] Y. 'Maulida, E. A. 'Desinaini, L. N. 'Utami, W. D. 'Yuliati, D. 'Farida, "Breast Cancer Survival Analysis Using Cox Proportional Hazard Regression and Kaplan Meier Method," *JTAM (Jurnal Teori dan Aplikasi Matematika)*, vol. 5, no. 2, pp. 340–358, Oct. 2021.
- [8] M. Herrmann, P. Probst, R. Hornung, V. Jurinovic, and A.-L. Boulesteix, "Large-scale benchmark study of survival prediction methods using multi-omics data," Mar. 2020, doi: [10.1093/bib/bbaa167](https://doi.org/10.1093/bib/bbaa167).
- [9] D. Chicco and G. Jurman, "Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, p. 16, Dec. 2020, doi: [10.1186/s12911-020-1023-5](https://doi.org/10.1186/s12911-020-1023-5).
- [10] A. R. Cuthbert, L. C. Giles, G. Glonek, L. M. Kalisch Ellett, and N. L. Pratt, "A comparison of survival models for prediction of eight-year revision risk following total knee and hip arthroplasty," *BMC Med. Res. Methodol.*, vol. 22, no. 1, p. 164, Jun. 2022, doi: [10.1186/s12874-022-01644-3](https://doi.org/10.1186/s12874-022-01644-3).
- [11] J. Katzman, U. Shaham, J. Bates, A. Cloninger, T. Jiang, and Y. Kluger, "DeepSurv: Personalized Treatment Recommender System Using A Cox Proportional Hazards Deep Neural Network," Aug. 2017, doi: [10.1186/s12874-018-0482-1](https://doi.org/10.1186/s12874-018-0482-1).
- [12] N. Asghar *et al.*, "Improved nonparametric survival prediction using CoxPH, Random Survival Forest & DeepHit Neural Network," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, p. 120, May 2024, doi: [10.1186/s12911-024-02525-z](https://doi.org/10.1186/s12911-024-02525-z).

- [13] P. Sankar and R. Bai V, “Survival Analysis on Relapse Predictions of Breast Cancer using Optimized ANN and DeepSurv Model,” in *2022 International Conference on Computing, Communication, Security and Intelligent Systems (IC3SIS)*, IEEE, Jun. 2022, pp. 1–6. doi: [10.1109/IC3SIS54991.2022.9885556](https://doi.org/10.1109/IC3SIS54991.2022.9885556).
- [14] L. M.K., S. N. S. Acharya, A. Kamath, and D. R. Micheal, “Mortality Prediction of Surgical Intensive Care Unit Patients Using Deep Learning Based Survival Models ,” *Malaysian Journal of Science*, vol. 41, no. 3, pp. 44–48, Oct. 2022, doi: [10.22452/mjs.vol41no3.6](https://doi.org/10.22452/mjs.vol41no3.6).
- [15] B. Yang *et al.*, “Development and Validation of a DeepSurv Nomogram to Predict Survival Outcomes and Guide Personalized Adjuvant Chemotherapy in Non-Small Cell Lung Cancer,” *Front. Oncol.*, vol. 12, Jun. 2022, doi: [10.3389/fonc.2022.895014](https://doi.org/10.3389/fonc.2022.895014).
- [16] C. Cui, Y. Tang, and W. Zhang, “Deep Contrastive Survival Analysis with Dual-View Clustering,” *Electronics (Basel)*, vol. 13, no. 24, p. 4866, Dec. 2024, doi: [10.3390/electronics13244866](https://doi.org/10.3390/electronics13244866).
- [17] X. Yang, H. Qiu, L. Wang, and X. Wang, “Predicting Colorectal Cancer Survival Using Time-to-Event Machine Learning: Retrospective Cohort Study,” *J. Med. Internet Res.*, vol. 25, p. e44417, Oct. 2023, doi: [10.2196/44417](https://doi.org/10.2196/44417).
- [18] Y. Liu, L. Xie, D. Wang, and K. Xia, “A deep learning algorithm with good prediction efficacy for cancer-specific survival in osteosarcoma: A retrospective study,” *PLoS One*, vol. 18, no. 9, p. e0286841, Sep. 2023, doi: [10.1371/journal.pone.0286841](https://doi.org/10.1371/journal.pone.0286841).
- [19] C. Jiang, K. Wang, L. Yan, H. Yao, H. Shi, and R. Lin, “Predicting the survival of patients with pancreatic neuroendocrine neoplasms using deep learning: A study based on Surveillance, Epidemiology, and End Results database,” *Cancer Med.*, vol. 12, no. 11, pp. 12413–12424, Jun. 2023, doi: [10.1002/cam4.5949](https://doi.org/10.1002/cam4.5949).
- [20] J. Zeng, D. Song, K. Li, F. Cao, and Y. Zheng, “Deep learning model for predicting postoperative survival of patients with gastric cancer,” *Front. Oncol.*, vol. 14, Apr. 2024, doi: [10.3389/fonc.2024.1329983](https://doi.org/10.3389/fonc.2024.1329983).
- [21] N. Bice *et al.*, “Deep learning-based survival analysis for brain metastasis patients with the national cancer database,” *J. Appl. Clin. Med. Phys.*, vol. 21, no. 9, pp. 187–192, Sep. 2020, doi: [10.1002/acm2.12995](https://doi.org/10.1002/acm2.12995).
- [22] J. Zeng, K. Li, F. Cao, and Y. Zheng, “Development and validation of survival prediction model for gastric adenocarcinoma patients using deep learning: A SEER-based study,” *Front. Oncol.*, vol. 13, Mar. 2023, doi: [10.3389/fonc.2023.1131859](https://doi.org/10.3389/fonc.2023.1131859).
- [23] Z. Wang *et al.*, “Survival modeling using deep learning, machine learning and statistical methods: A comparative analysis for predicting mortality after hospital admission,” Mar. 2024.
- [24] T. G. Baidoo and H. Rodrigo, “Data-driven survival modeling for breast cancer prognostics: A comparative study with machine learning and traditional survival modeling methods,” *PLoS One*, vol. 20, no. 4, p. e0318167, Apr. 2025, doi: [10.1371/journal.pone.0318167](https://doi.org/10.1371/journal.pone.0318167).
- [25] N. E. Alsubaie, F. El Guma, A. G. M. Musa, H. A. Abdali, M. Awadalla, and B. Alosaimi, “Comparative evaluation of Cox regression and machine learning approaches for survival prediction in stage IV lung cancer,” *Discover Computing*, vol. 29, no. 1, p. 102, Feb. 2026, doi: [10.1007/s10791-026-10014-2](https://doi.org/10.1007/s10791-026-10014-2).
- [26] Di Asih I Maruddani, Tarno, Abdul Hoyyi, Rita Rahmawati, and Yuciana Wilandari, *Survival Analysis*. Universitas Diponegoro, 2021.
- [27] S. V. Deo, V. Deo, and V. Sundaram, “Survival analysis—part 2: Cox proportional hazards model,” *Indian J. Thorac. Cardiovasc. Surg.*, vol. 37, no. 2, pp. 229–233, Mar. 2021, doi: [10.1007/s12055-020-01108-7](https://doi.org/10.1007/s12055-020-01108-7).
- [28] C. P. Obite, E. O. Chukwudi, M. Uchechukwu, and U. I. Nwosu, “Factor enhanced DeepSurv: A deep learning approach for predicting survival probabilities in cirrhosis data,” *Comput. Biol. Med.*, vol. 189, p. 109963, May 2025, doi: [10.1016/j.compbiomed.2025.109963](https://doi.org/10.1016/j.compbiomed.2025.109963).

- [29] E. Zhu *et al.*, “Individualized Survival Prediction and Surgery Recommendation for Patients with Glioblastoma,” *Front. Med. (Lausanne)*, vol. 11, May 2024, doi: [10.3389/fmed.2024.1330907](https://doi.org/10.3389/fmed.2024.1330907).
- [30] N. Bozhbanbayeva *et al.*, “DeepSurv: A machine learning model for predicting mortality in very-low-birth-weight infants treated in intensive care units,” Aug. 11, 2025. doi: [10.21203/rs.3.rs-7126001/v1](https://doi.org/10.21203/rs.3.rs-7126001/v1).
- [31] S. Malik *et al.*, “A unified multi modal transformer framework for breast cancer recurrence prediction and survival analysis,” *Sci. Rep.*, vol. 16, no. 1, p. 8334, Feb. 2026, doi: [10.1038/s41598-026-37046-4](https://doi.org/10.1038/s41598-026-37046-4).