IJSA (2025), Vol. 9, No. 1, 33–45
DOI: https://doi.org/10.29244/ijsa.v9i1p33-44

Indonesian Journal of Statistics and Its Applications
Available at https://journal-stats.ipb.ac.id/index.php/ijsa
e-ISSN: 2599-0802

Early Preeclampsia Detection Using XGBoost-Cox Proportional Hazard Model*

Arya Wira Syahdwinata[1], Sarini Abdullah[1‡]

[1,1]Department of Mathematics, University of Indonesia, Indonesia
[‡]corresponding author: sarini@sci.ui.ac.id

Abstract

Various prognostic models based on survival analysis methods have been proposed to predict the risk of preeclampsia (PE). To develop a more accurate yet interpretable prediction approach, we utilized clinical data from pregnant women collected at a hospital in Jakarta and applied the XGBoost-Cox Proportional Hazard Model (XGB-Cox). This model integrates the predictive power of the XGBoost machine learning algorithm with the Cox Proportional Hazard (Cox-PH) model, which estimates the effect of covariates on event time. Our results show that the XGB-Cox model outperforms the traditional Cox-PH model based on four evaluation metrics: log-likelihood, log-rank test, concordance index (C-index), and Brier score. The XGB-Cox model achieved a higher C-index of 0.8908 compared to 0.7548 for Cox-PH, indicating improved risk discrimination. Kaplan-Meier curves suggest that XGB-Cox provides better separation across risk quartiles. While XGB-Cox generally yields lower Brier Scores, its performance declines at later gestational weeks. The Cox-PH model remains superior in interpretability, offering clear hazard ratios, while XGB-Cox enhances model fitness and still provides meaningful insights into feature importance. Additionally, sensitivity analysis underscores the need to carefully determine the proportion of censored data, as excessive censoring affects model stability. These findings suggest that XGB-Cox provides a robust predictive framework for early PE risk assessment, supporting its potential application in clinical decision-making for maternal healthcare.

Keywords: Cox Proportional Hazard, Preeclampsia, Risk Prediction, Survival Analysis, XGBoost.

*Received: May 2025; Reviewed: Jun 2025; Published: Jun 2025

1 Background

Preeclampsia (PE) poses a serious threat to the health of pregnant women and unborn children, potentially leading to complications such as preterm birth, placental abruption, and intrauterine growth restriction (American College of Obstetricians and Gynecologists, 2019). It also increases the long-term risk of cardiovascular disease for both mother and child (Sibai, 2012). Despite advances in obstetric care, PE remains one of the leading causes of maternal and perinatal morbidity and mortality (Roberts & Hubel, 2013). Globally, it affects approximately 2–8% of pregnancies (American College of Obstetricians and Gynecologists, 2020), with an estimated 12% of maternal deaths attributable to PE and its complications (World Health Organization, 2024). In Southeast Asia, Indonesia ranks fourth in maternal mortality, with 173 maternal deaths per 100,000 live births in 2020 (United Nation, 2023). This rate is significantly higher than that of neighboring countries such as Brunei, Thailand, Malaysia, and Singapore. Between 2016 and 2022, approximately 18.5% of maternal deaths in Indonesia were caused by PE (Syairaji *et al.*, 2024), making it one of the most prominent contributors to maternal mortality in the country.

Identifying pregnancies at high risk of developing PE is crucial, as timely therapeutic interventions including prophylactic aspirin use, closer monitoring, and earlier delivery can reduce the incidence and severity of PE and its associated complications (Wright *et al.*, 2020). With the advancement of computational tools and methods, various techniques have been developed to improve early detection and management of PE. For instance, machine learning algorithms can analyze PE datasets to identify patterns and predict the likelihood of PE with higher accuracy. Classification algorithms such as Random Forest (RF), Linear Regression (LR), Support Vector Machines (SVM), and Generalized Linear Models (GLM) have been used for early PE prognosis using gene expression data (Sharma *et al.*, 2023). However, these approaches only predict whether a case will result in PE or not, without estimating risk probabilities. Moreover, insights drawn from genomic data are often difficult to interpret. Logistic regression models have also been applied to predict PE probability and identify risk factors (Fang *et al.*, 2024), but they lack the temporal dimension needed to understand disease progression over time.

Survival analysis techniques have emerged as powerful approaches for PE prediction. By leveraging time-to-event data, these methods estimate the probability of developing PE within a specified timeframe, allowing healthcare providers to implement timely and targeted interventions (Kleinbaum & Klein, 2012). Bayesian competing risk models (Devana & Abdullah, 2022) offer advantages over conventional survival models, which typically assume that women who deliver without developing PE are censored and may still experience the condition post-gestation. In contrast, competing risk models simultaneously account for multiple possible outcomes, offering a more accurate picture of morbidity. Nevertheless, such models have limitations, including their reliance on distributional assumptions (e.g., Weibull) that may not capture real-world data complexity. Moreover, their Bayesian formulation can be sensitive to prior selection, especially when the sample size is small (Lindley, 1957; Naaman, 2016).

Recent studies have shown that Extreme Gradient Boosting (XGBoost) outperforms deep learning models in classification and regression tasks involving tabular data (Shwartz-Ziv & Armon, 2022). XGBoost not only achieves high accuracy

but also demands significantly lower computational resources, making it a more efficient option. This is particularly relevant given that PE-related data is often tabular in nature. Additionally, XGBoost offers built-in feature importance metrics, providing interpretable insights into factors contributing to PE. A hybrid approach called EXSA, which combines XGBoost with the Cox Proportional Hazard (Cox-PH) model, has been used to predict log-hazard ratios in breast cancer data (Liu *et al.*, 2021). EXSA captures nonlinear relationships without requiring prior assumptions about time-to-event distributions or prior probabilities, offering a flexible alternative to traditional models.

While the above methods focus on achieving the best possible model fit, selecting the right amount of censored data presents an additional challenge. A recent study highlights how increasing censoring can degrade prediction accuracy, making survival estimates less reliable (Báskay *et al.*, 2025). This underscores the need for careful censoring strategies to maintain robust survival predictions.

This study aims to develop a predictive model for early detection of PE using the XGB-Cox approach, which integrates the Cox-PH model with the XGBoost algorithm. By combining the time-awareness of Cox-PH with the predictive power and interpretability of XGBoost, this research explores whether the hybrid method can address limitations of conventional survival analysis and support more timely and effective clinical interventions for maternal healthcare. It also examines how different levels of censoring affect model performance, helping to determine the optimal balance for improved predictive accuracy.

2 Methodology

2.1 Data

The data used is secondary data from a screening of pregnant women in their first trimester who gave birth either with or without PE at a hospital in Jakarta between August 2019 and October 2020.

2.2 Cox Proportional Hazard

The Cox Proportional Hazards (Cox-PH) model is one of the most commonly used methods in survival analysis. This model assumes that the hazard ratio between two individuals remains constant over time. The formula is expressed as:

$$h(t, X) = h_0(t) e^{\sum_{i=1}^{p} \beta_i X_i} \tag{1}$$

where $h(t, X)$ is the hazard function at time t, $h_0(t)$ is the baseline hazard function, X is the vector of predictor variables, and β represents the coefficients to be estimated (Cox, 1972).

In the Cox-PH model, the likelihood is calculated as a partial likelihood, which avoids specifying the baseline hazard distribution $h_0(t)$. The partial likelihood is given by:

$$L(\beta) = \prod_{i \in D} \frac{e^{\sum_{k=1}^{p} \beta_k x_{ik}}}{\sum_{j \in R(t_i)} e^{\sum_{k=1}^{p} \beta_k x_{jk}}} \tag{2}$$

Where:

- D is the set of individuals who experienced the event at time t_i,
- $R(t_i)$ is the risk set—individuals still at risk just before time t_i,
- β are the regression coefficients of the predictor variables,
- x_i represents the predictor variables for individual i.

From the partial likelihood, the log-likelihood function can be derived as follows:

$$\log L(\beta) = \sum_{i \in D} \left(\sum_{k=1}^{p} \beta_k x_{ik} - \log \sum_{j \in R(t_i)} e^{\sum_{k=1}^{p} \beta_k x_{jk}} \right) \tag{3}$$

In the Cox-PH model, the estimation of β is obtained by maximizing the log-likelihood function (Cox, 1972).

2.3 **Extreme Gradient Boosting**

XGBoost is a powerful machine learning algorithm used for both regression and classification problems. It employs a boosting method that incrementally builds models by optimizing an objective function, typically using Mean Squared Error (MSE) as the loss criterion (Chen & Guestrin, 2016). The general objective function of XGBoost is:

$$L = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{j=1}^{k} \Omega(f_j) \tag{4}$$

Where l is the loss function, $\hat{y}_i$ is the prediction, and $\Omega(f_j)$ is the regularization term for the jth model to prevent overfitting (Chen & Guestrin, 2016).

XGBoost uses gradient boosting to iteratively build weak decision tree models and refine predictions until optimal performance is achieved. At each stage, a new tree is built to minimize the residuals of the previous tree, where the residual is the difference between the actual value (y_i) and the predicted value ($\hat{y}_i$).

XGBoost is widely used in regression problems due to its ability to produce accurate models with efficient computation. In regression tasks, XGBoost typically uses Mean Squared Error (MSE) as the objective function to minimize prediction errors while balancing model complexity.

Data splitting in XGBoost is guided by the Gain criterion, defined as:

$$\text{Gain} = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \tag{5}$$

Where:

- G is the gradient (first derivative of the loss function),
- H is the Hessian (second derivative of the loss function),
- λ and γ are regularization parameters to prevent overfitting (Chen & Guestrin, 2016).

XGBoost can be adapted for survival analysis by integrating it with the Cox Proportional Hazards (Cox-PH) model. In this context, the goal is to predict the log-hazard ratio rather than directly predicting the target variable (time and event status).

To implement this, the loss function in Equation (4) is replaced with the log-likelihood function (3) of the Cox-PH model. The negative log-likelihood is then minimized, which is equivalent to maximizing the Cox partial log-likelihood as the objective.

The splitting criterion in Equation (5) is applied with gradients G and Hessians H derived from the Cox partial log-likelihood (2), instead of using MSE.

This approach provides greater flexibility than the conventional Cox-PH model, which assumes a constant hazard ratio. By integrating XGBoost, the model can capture complex nonlinear relationships among features, leading to more accurate survival predictions (Liu *et al.*, 2021).

2.4 **Performance Metrics for Survival Models**

To evaluate the performance of survival models, several common metrics are used to compare model prediction outcomes. The following are widely used metrics along with the formulas that underpin them.

The log-rank test is a non-parametric statistical test used to compare survival curves between two or more groups. It evaluates the null hypothesis that there is no difference in survival functions between the groups. The log-rank statistic is calculated using the following formula:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \tag{6}$$

where:

- O_i is the observed number of events in group i,
- E_i is the expected number of events based on the combined survival function.

The log-rank test produces a p-value used to determine whether the difference between groups is statistically significant (Mantel, 1966).

The concordance index (C-index) is a performance metric used to assess the accuracy of time-to-event predictions in survival analysis (Harrell, 1982). It is defined as:

$$C = \frac{\sum_{i,j} I(\tilde{T}_i > \tilde{T}_j) \cdot I(\eta_j > \eta_i)}{\sum_{i,j} I(\tilde{T}_i > \tilde{T}_j)} \tag{7}$$

where:

- $I(\tilde{T}_i > \tilde{T}_j)$ is an indicator function that equals 1 if the event time $\tilde{T}_i$ is greater than $\tilde{T}_j$, and 0 otherwise,
- $I(\eta_j > \eta_i)$ is an indicator that equals 1 if the predicted risk score η_j is greater than η_i, and 0 otherwise.

The C-index evaluates how well the survival model ranks individuals based on predicted risk scores. It measures the proportion of subject pairs that are correctly ordered by the model, such that a higher predicted risk corresponds to a shorter survival time. The C-index ranges from 0 to 1, where 0 indicates completely incorrect predictions, 0.5 suggests performance no better than random guessing, and 1 indicates perfect prediction (Harrell, 1982).

The Brier score is a metric that quantifies the prediction error in survival data by incorporating the cumulative probability of the predicted event. It is defined as:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2 \tag{8}$$

Where:

- N is the number of observations,
- f_i is the predicted probability for the i-th event,
- o_i is the actual outcome (1 if the event occurred, 0 otherwise).

The Brier score measures how well the model combines predictive accuracy for both event occurrence and censored data. Lower scores indicate better prediction performance (Brier, 1950).

3 Result and Discussion

3.1 Data Preprocessing

The data used in this study are comprising a total of 946 observations, with 71 experiencing PE and 875 not. It includes various clinical (e.g., `PreviousgestationalHT`, `DiabetesMellitusType2`, `UseofAspirin`), demographic (e.g., `class_Age`, `class_BMI`, `AnyFamilyhistoryofPE`), and biochemical (e.g., `FinalMAP`, `FinalMeanUtAPI`, `PLGFconcentration_pgml`) variables gathered during pregnancy.

Binary predictors were converted to categorical format. Numerical predictors were retained in their original scales to preserve clinical interpretability. All values in the dataset were complete and no imputation was necessary.

3.2 Model Development

All analyses were conducted using the R programming language version 4.5.0. Two models were developed: a Cox-PH model implemented using the `coxph()` function from the `survival` package (Therneau, 2024), and a XGB-Cox with the `xgb.train()` function with `survival:cox` objective from `xgboost` package (Chen *et al.*, 2025).

To ensure balanced data for survival analysis and evaluate the impact of censoring on model performance, a sensitivity analysis was conducted by simulating both XGB-Cox and Cox-PH models across different levels of censored data. The simulation ran 100 iterations for each censoring percentage, ranging from 0% to 90%, and assessed performance through log-likelihood metrics. In each iteration, censored

data points were randomly selected across the dataset to ensure variability. The censoring percentage with the smallest mean log-likelihood difference between models was chosen to optimize stability and predictive accuracy.

Hyperparameter tuning on XGB-Cox performed using random search with 1000 rounds, each evaluated through 10-fold cross-validation to prevent overfitting. Tuned parameters include `eta` (learning rate), `max_depth` (tree depth), `gamma` (split threshold), `colsample_bytree` (feature sampling), `min_child_weight` (node weight limit), `subsample` (data fraction per round), `lambda` (L2 regularization), and `alpha` (L1 regularization).

3.3 **Model Performance Evaluation**

This study evaluates the predictive performance of the survival models using four metrics: Log-likelihood, Log-rank, C-index, and Brier Score. Each metric provides complementary insights into the model's ability to discriminate risk, rank patients appropriately, and accurately estimate event probabilities over time.

To determine the optimal balance between censored and uncensored data, a sensitivity analysis was conducted. Results revealed that log-likelihood worsens as the proportion of censored data increases for both models. XGB-Cox initially performs slightly better than Cox-PH up to 55% censoring, but beyond this point, its log-likelihood declines more rapidly, making it less stable under high censoring conditions. This suggests that XGB-Cox is more sensitive to added censored data, affecting its predictive reliability. The difference in mean log-likelihood between the two models was minimized at 55% censoring, making this threshold the most appropriate choice for fair model comparison and stability. Based on this finding, all subsequent experiments will be conducted using randomly selected censored data at this threshold.

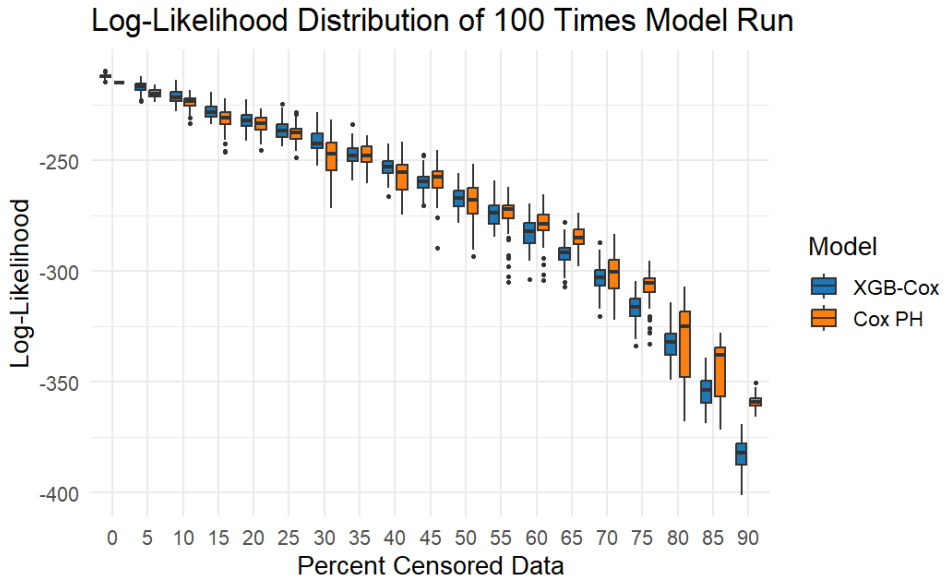

Figure 1: Log-likelihood distribution of 100 times run on various censored data percentage.

Using data subset acquired, hyperparameter tuning was performed to refine the

XGB-Cox model for improvement. Once tuning was completed, risk stratification was assessed by generating Kaplan-Meier (KM) curves based on quartiles of the predicted risk scores from both Cox-PH and XGB-Cox models.

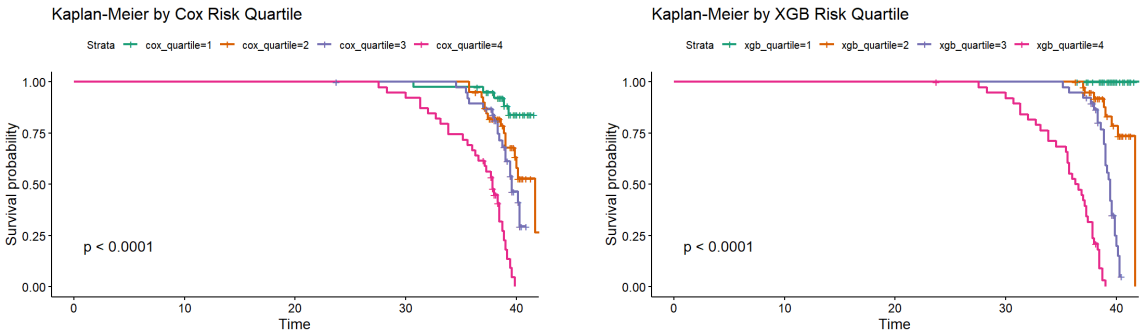

Figure 2: Kaplan-Meier Curves for Risk Quartiles for Cox-PH (left) and XGB-Cox (right).

Although the log-rank test p-values for both models were less than 0.0001, indicating statistically significant differences for at least one risk group, the Cox-PH model fails to clearly distinguish between the its intermediate risk quartiles, as their KM curves overlap. In contrast, the XGB-Cox model achieves better separation across all quartiles, suggesting improved stratification of patients by predicted risk.

To evaluate the model's ability to rank patients by risk, the C-index was computed. The C-index reflects the model's capacity to assign higher risk scores to individuals who experience the event earlier.

Table 1: C-index Comparison Between Models

Model	**C-index**
Cox-PH	0.7547
XGB-Cox	0.8907

The results show that the XGB-Cox model outperforms the Cox-PH model in ranking individuals by their time-to-event. A higher C-index indicates that XGB-Cox more accurately assigns higher risk scores to patients who develop PE earlier, which is essential for early detection.

However, while the C-index measures discriminatory ability, it does not evaluate how well the predicted probabilities align with actual outcomes. To assess the overall accuracy and calibration of probabilistic predictions, the Brier Score was calculated at various gestational time points.

The selected time points correspond to quantiles of the observed event times, focusing on later gestational stages. This choice reflects the clinical reality that PE events tend to occur closer to delivery. Evaluating the models at these critical time windows ensures relevance to real-world decision-making and highlights predictive performance when accurate forecasting is most needed.

Across most time points, the XGB-Cox model yields lower Brier Scores than the Cox-PH model, indicating better alignment between predicted risk probabilities and actual outcomes. However, at later gestational weeks, the Brier Score for XGB-Cox increases and surpasses that of Cox-PH, suggesting a decline in predictive accuracy as gestational age progresses. This shift implies that while XGB-Cox performs well in earlier predictions, its reliability diminishes in later stages.

Table 2: Brier Score Comparison at Different Time Points

Time (Weeks of Gestation)	Brier Score (Cox-PH)	Brier Score (XGB-Cox)
32.71	0.9322	0.3551
35.57	0.8361	0.2407
36.57	0.7772	0.2242
37.29	0.6941	0.2308
38.00	0.6341	0.2686
38.43	0.6047	0.3091
38.86	0.5535	0.3589
39.29	0.5134	0.4253
39.86	0.5095	0.5215

3.4 **Model Interpretation**

The Cox-PH model provides coefficients representing the log hazard ratios for each feature, making it highly interpretable. A positive coefficient indicates an increased risk of PE, while a negative coefficient suggests a protective effect. The coefficients show that variables like `DiabetesMellitusType2`, `Conception`, `UseofAspirin`, `PreviousPE`, `AnyFamilyhistoryofPE`, and `FinalMeanUtAPI` contribute to increased PE risk, while `class_Age1`, `UseofantiHTdrug` and `ChronicHT` have negative coefficients, indicating a potential protective effect.

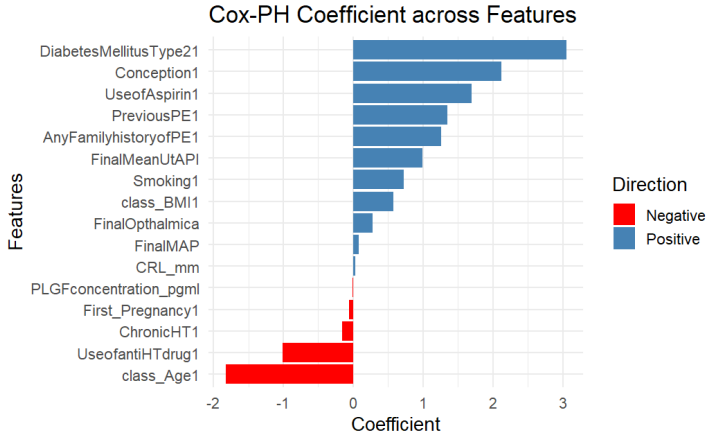

Figure 3: Coefficients from the Cox-PH model representing log hazard ratios of features.

While Cox-PH excels in interpretability by providing direct hazard ratios, the XGB-Cox model enhances predictive accuracy while still retaining some interpretability. Feature importance in XGB-Cox is measured using gain, which quantifies how much each feature improves the model's predictive accuracy when used for splitting in decision trees. The most influential features in XGB-Cox are `FinalMAP`, `FinalOpthalmica`, and `FinalMeanUtAPI`, followed by `PLGFconcentration_pgml` and `CRL_mm`. `PreviousPE1` has comparatively lower importance, indicating a smaller contribution to the model's predictions.

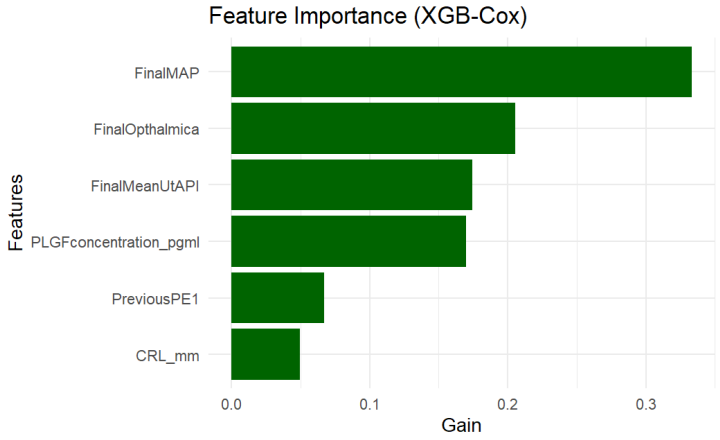

Figure 4: Feature importance in the XGB-Cox model based on gain.

3.5 **Discussion**

A trade-off exists between interpretability and predictive performance in survival modeling. The Cox-PH model provides clearer insight into how individual predictors affect the risk of PE, making it preferable when transparency is essential. In contrast, the XGB-Cox model enhances predictive accuracy while retaining interpretability through feature importance scores, making it an alternative for applications where prediction takes priority.

An additional consideration is the model's sensitivity to the level of censoring in the data. By varying the proportion of censored observations and evaluating changes in log-likelihood, it was observed that increased censoring generally reduced model fit. This outcome suggests that survival models, particularly XGB-Cox may struggle to capture event-time patterns when too much of the data is censored. Therefore, careful attention to the degree of censoring is essential to ensure robust model development and reliable performance evaluation.

4 **Conclusion**

This study compared the predictive performance and interpretability of the Cox Proportional Hazards (Cox-PH) model and the XGBoost-based Cox model (XGB-Cox) for early detection of PE using clinical, demographic, and biochemical data from 946 pregnant patients.

The XGB-Cox model demonstrated better performance across multiple evaluation metrics. It achieved a higher C-index of 0.7547 compared to 0.8907 for the Cox-PH model, indicating better discrimination in ranking patients by risk of PE. Kaplan-Meier curves further revealed that the XGB-Cox model provided clearer stratification of risk quartiles, with well-separated survival curves, unlike the Cox-PH model where some risk groups overlapped.

In terms of calibration, the XGB-Cox model generally yielded lower Brier Scores at multiple gestational time points, reflecting improved overall alignment between predicted and actual risk probabilities. However, at later stages, its Brier Score surpassed that of Cox-PH, suggesting a decline in predictive accuracy as gestational age progresses.

Interpretation of the models revealed that the Cox-PH coefficients highlight features with positive or negative effects on hazard, making it superior in interpretability. The most relevant predictors included `DiabetesMellitusType2`, `Conception`, `UseofAspirin`, `PreviousPE`, `class_Age`, `UseofantiHTdrug` and `ChronicHT`. Meanwhile, the XGB-Cox model enhances predictive accuracy while still offering interpretability through feature importance scores, ranking `FinalMAP`, `FinalOpthalmica`, and `FinalMeanUtAPI` as the most influential variables.

Furthermore, the study highlights the significance of carefully considering the proportion of censored data included in survival analysis. Our sensitivity analysis demonstrated that increasing censoring beyond 55% negatively impacts model stability, particularly in XGB-Cox, which is more sensitive to censored observations. This finding underscores the need for a balanced approach when selecting censored data to ensure fair comparisons and reliable predictions.

Overall, the results suggest that while Cox-PH remains preferable for interpretability, XGB-Cox provides a more accurate predictive approach while retaining meaningful insights into feature importance. This balance between fitness and interpretability highlights its potential utility in clinical decision making for maternal care. In practice, doctors could use a prediction tool based on the XGB-Cox model, where routinely collected data during pregnancy check-ups can be entered for an individual patient to estimate their future risk of developing PE, helping them assess risk progression over time and support timely interventions.

References

American College of Obstetricians and Gynecologists (2019). Practice bulletin no. 202: Gestational hypertension and preeclampsia. *Obstet Gynecol*, *133*(1): 1–25. doi:10.1097/AOG.0000000000003018.

American College of Obstetricians and Gynecologists (2020). Practice bulletin no. 222: Gestational hypertension and preeclampsia. *Obstet Gynecol*, *135*(6): e237–e260. doi:10.1097/AOG.0000000000003891.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*(1): 1–3. doi:10.1175/1520-0493(1950)078<0001:VOFETI>2.0.CO;2.

Báskay, J., Mezei, T., Banczerowski, P., Horváth, A., Joó, T., & Pollner, P. (2025). Censoring sensitivity analysis for benchmarking survival machine learning methods. *Sci*, *7*(1). ISSN 2413-4155. doi:10.3390/sci7010018.

Chen, T. & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794. doi:10.1145/2939672.2939785.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Geng, Y., Li, Y., Yuan, J., & Cortes, D. (2025). *xgboost: Extreme Gradient Boosting*. R package version 3.1.0.0.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2): 187–220.

Devana, N. & Abdullah, S. (2022). Bayesian competing risk model for medical data. *Proceedings of the International Conference on Science and Engineering (ICSE-UIN-SUKA 2021)*.

Fang, Y., Liu, H., Li, Y., Cheng, J., Wang, X., Shen, B., Chen, H., & Wang, Q. (2024). A prediction model of preeclampsia in hyperglycemia pregnancy. *Diabetes, Metabolic Syndrome and Obesity*, *17*: 1321–1333. doi:10.2147/dmso.s453204.

Harrell, F. E. (1982). Evaluating the yield of medical tests. *The Journal of the American Medical Association*, *247*(18): 2543. doi:10.1001/jama.1982.03320430047030.

Kleinbaum, D. G. & Klein, M. (2012). *Survival Analysis: A Self-Learning Text, Third Edition*. Springer Science & Business Media.

Lindley, D. V. (1957). A statistical paradox. *Biometrika*, *44*: 187–192.

Liu, P., Fu, B., Yang, S. X., Deng, L., Zhong, X., & Zheng, H. (2021). Optimizing survival analysis of xgboost for ties to predict disease progression of breast cancer. *IEEE transactions on biomedical engineering*, *68*(1): 148–160. doi:10.1109/TBME.2020.2993278.

Mantel, N. (1966). Evaluation of survival data and two-sample tests. *Journal of the American Statistical Association*, *61*(213): 1096–1120. doi:10.1080/01621459.1966.10482521.

Naaman, M. (2016). Almost sure hypothesis testing and a resolution of the Jeffreys-Lindley paradox. *Electronic Journal of Statistics*, *10*(1): 1526 – 1550. doi:10.1214/16-EJS1146.

Roberts, J. M. & Hubel, C. A. (2013). The two stage model of preeclampsia: variations on the theme. *Placenta*, *34*(5): 297–303. doi:10.1016/j.placenta.2013.02.009.

Sharma, P., Pandey, S., Ray, S., Mazumder, M., Gupta, P., Sengupta, A., Chaurasia, A., & Sengupta, A. (2023). Predictive modeling of recurrent implantation failure and pre-eclampsia using machine learning and gene expression profiling. In *International Conference on Computer, Communication, and Signal Processing*. doi:10.1109/icccnt56998.2023.10307984.

Shwartz-Ziv, R. & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion*, *81*: 84–90. ISSN 1566-2535. doi:https://doi.org/10.1016/j.inffus.2021.11.011.

Sibai, B. M. (2012). Etiology and management of preeclampsia. *American Journal of Obstetrics and Gynecology*, *206*(6): 470–479. doi:10.1016/j.ajog.2012.03.002.

Syairaji, M., Detty, S., Nurdiati, Bayu, S., Wiratama, Zita, D., Prüst, Kitty, W. M., Bloemenkamp, Kim, J. C., & Verschueren (2024). Trends and causes of maternal mortality in indonesia: a systematic review. *BMC Pregnancy and Childbirth*, *24*(1). doi:10.1186/s12884-024-06687-6.

Therneau, T. M. (2024). *A Package for Survival Analysis in R*. R package version 3.8-3.

United Nation (2023). *Trends in maternal mortality 2000 to 2020: estimates by WHO, UNICEF, UNFPA, World Bank Group and UNDESA/Population Division*. World Health Organization.

World Health Organization (2024). Maternal mortality.

Wright, D., Wright, A., & Nicolaides, K. H. (2020). The competing risk approach for prediction of preeclampsia. *American Journal of Obstetrics and Gynecology*, *223*(1): 12–23.e7. doi:10.1016/j.ajog.2019.11.1247.