Scientific Journal of Informatics
Vol. 13, No. 3, August 2026

p-ISSN 2407-7658 http://journal.unnes.ac.id/nju/index.php/sji e-ISSN 2460-0040

Comparison of Tree-Based Survival Models for Predicting Graduate Study Duration and Risk Stratification

Ahmad Syauqi[1*], Anwar Fitrianto[2], Hari Wijayanto[3]
[1, 2, 3]Statistics and Data Science Study Program, IPB University, Indonesia

Abstract.
Purpose: This study aims to develop and evaluate a tree-based survival machine learning framework for predicting graduate study duration and identifying students at risk of delayed graduation. The study addresses the growing need for accurate educational time-to-event prediction to support academic monitoring and data-driven decision-making in higher education institutions, including IPB University.
Methods: A quantitative predictive analysis was conducted using data from 3,417 master students from the 2020–2022 cohorts. Four tree-based survival models were evaluated, namely Survival Tree (ST), Extremely Randomized Survival Tree (EST), Random Survival Forest (RSF), and Gradient Boosting Survival (GBS). The analysis used right-censored survival data with a 42-month observation period. Model evaluation was conducted using repeated stratified random split validation (10 repetitions) with Concordance Index (C-index) and Integrated Brier Score (IBS) metrics. Risk stratification was subsequently performed using the best-performing model based on predicted survival probabilities at a 24-month time horizon.
Result: GBS achieved the best overall predictive performance with the highest mean C-index (0.659) and the lowest mean IBS (0.194), indicating superior discrimination and prediction accuracy compared to ST, EST, and RSF. The repeated evaluation results also demonstrated stable predictive performance across data partitions. Risk stratification successfully separated students into low-, medium-, and high-risk groups with significantly different survival patterns. High-risk students generally tended to be older, have lower undergraduate GPA, were more often male, and more frequently originate from private undergraduate institutions.
Novelty: This study provides a comparative evaluation of multiple tree-based survival machine learning models within an educational time-to-event framework. The integration of repeated survival model evaluation with practical student risk stratification offers both methodological and applied contributions for academic monitoring and early intervention strategies in higher education.

Keywords: Survival analysis, Machine learning, Gradient boosting survival, Random survival forest, Risk stratification
Received May 2026 / **Revised** July 2026 / **Accepted** July 2026

This work is licensed under a Creative Commons Attribution 4.0 International License.

INTRODUCTION

The duration of graduate study is an important indicator of academic performance and institutional effectiveness in higher education. Timely graduation reflects the quality of academic systems, student support services, and educational management, while delayed graduation may negatively affect both students and institutions. In many universities, including IPB University, prolonged study duration remains a significant concern that requires more effective monitoring and prediction strategies. Consequently, educational data mining and predictive analytics have been increasingly utilized to support student retention, academic planning, and institutional decision-making [1], [2]. Similar data-driven applications have also been implemented in broader educational system development [3].

Recent advances in educational data mining and machine learning have substantially improved predictive modeling for academic monitoring and institutional decision-making in higher education [4], [5]. The predictors used in educational prediction studies also vary depending on the research objectives and available data. Previous studies have commonly employed demographic, academic, and behavioral attributes, such as age, gender, academic performance, and learning management system interactions, to

*Corresponding author.
Email addresses: sokinter5@gmail.com (Syauqi)*, anwarstat@gmail.com (Fitrianto), hari@apps.ipb.ac.id (Wijayanto)
DOI: 10.15294/sji.v13i3.48890

predict student success, retention, or academic performance [6], [7], [8]. These diverse predictor characteristics indicate that educational data often comprise heterogeneous numerical and categorical attributes, requiring predictive models capable of capturing complex relationships among variables.

However, predicting graduate study duration differs from conventional classification tasks because it represents a longitudinal time-to-event outcome that commonly includes right-censored observations and heterogeneous academic and demographic characteristics. Therefore, survival analysis has emerged as a suitable framework for educational time-to-event prediction because it explicitly models the time until an event occurs while accommodating censored observations [9], [10]. Traditional survival models such as the Cox proportional hazards model have been widely applied in time-to-event studies; however, these approaches rely on assumptions that may limit their ability to capture nonlinear relationships and complex interactions within educational datasets. Recent developments in machine learning-based survival analysis provide more flexible alternatives for modeling nonlinear relationships and heterogeneous data structures [11]. In addition, the integration of machine learning and survival analysis has shown promising performance across various predictive applications, including healthcare and risk prediction tasks [12], [13]. These developments indicate the increasing relevance of survival machine learning approaches in longitudinal educational analysis.

Among survival machine learning methods, tree-based approaches have received considerable attention due to their flexibility, interpretability, and ability to model nonlinear interactions without strong parametric assumptions. These tree-based approaches have been extensively investigated in healthcare, finance, and other survival prediction problems because they effectively capture nonlinear relationships and complex feature interactions while accommodating censored observations. Survival Trees (ST) provide a nonparametric framework for partitioning survival data into homogeneous groups, while ensemble-based methods such as Random Survival Forest (RSF) and Gradient Boosting Survival (GBS) improve predictive performance through aggregation and iterative learning strategies [14], [15]. Large-scale benchmarking studies and comparative evaluations consistently report that ensemble-based survival models, particularly RSF and boosting-based approaches, generally outperform single-tree models in terms of predictive discrimination and robustness across diverse survival datasets [16], [17]. Recent developments in boosting-based and scalable survival frameworks further strengthen the capability of advanced machine learning models in structured time-to-event prediction [18], [19]. Nevertheless, most comparative evaluations have been conducted using medical or financial datasets, while evidence from educational time-to-event prediction remains limited. Furthermore, model performance is highly dependent on dataset characteristics and evaluation strategies, emphasizing the need for comprehensive comparative assessment in educational settings.

Despite the growing literature on survival machine learning, several limitations remain in existing studies. Most previous studies focus on healthcare or medical datasets, while educational applications remain relatively limited. Moreover, limited studies have systematically compared multiple tree-based survival models, including ST, EST, RSF, and GBS, within a unified evaluation framework using real educational survival data. Although Extremely Randomized Survival Trees (EST) has demonstrated promising computational efficiency in simulation studies, its application to real educational survival datasets remains limited [17]. Furthermore, many predictive studies primarily emphasize model performance without integrating risk stratification for practical academic monitoring and intervention. These limitations indicate the need for more comprehensive educational survival modeling frameworks that not only compares multiple tree-based survival models but also integrates robust model evaluation with practical risk stratification to support data-driven academic decision-making.

Therefore, this study aims to develop and evaluate a tree-based survival machine learning framework for predicting graduate study duration among master students at IPB University. The study compares the performance of ST, EST, RSF, and GBS using repeated stratified random split evaluation with Concordance Index (C-index) and Integrated Brier Score (IBS) metrics. In addition, the best-performing model is further utilized for student risk stratification based on delayed graduation probability. The proposed framework integrates comparative survival machine learning evaluation with practical educational risk profiling, providing both predictive insights and decision-support relevance for academic monitoring in higher education.

METHODS

This study employs a quantitative approach using survival analysis combined with machine learning techniques to predict graduate study duration and perform student risk stratification. The methodology includes data preparation, model development, model evaluation, and risk stratification. Survival analysis was selected because it is suitable for modeling educational time-to-event data while accommodating censored observations. Several tree-based survival machine learning models were implemented and comparatively evaluated to identify the most reliable predictive framework. The overall research workflow is illustrated in Figure 1, which summarizes the sequential stages of data preparation, model development, performance evaluation, and risk stratification.

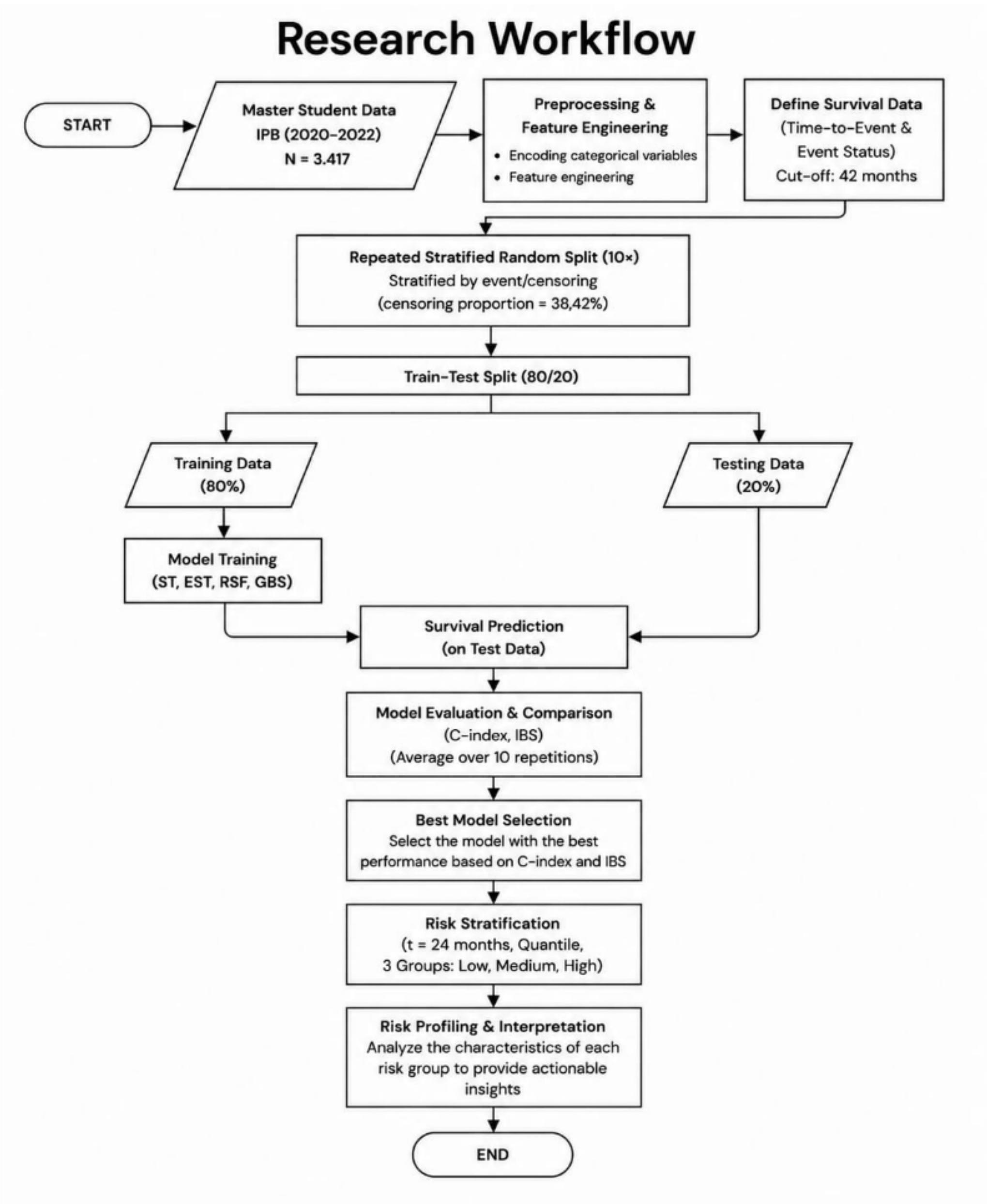

Figure 1. Research workflow

As shown in Figure 1, the proposed framework consists of four main stages: data preparation, model development, model evaluation, and risk stratification. Each stage is described in the following subsections. The data used in this study were obtained from the Directorate of Educational Administration of IPB University. The dataset consists of 3,417 master student records from cohorts 2020 to 2022. The observed data are right-censored, indicating that some students had not yet experienced the event of interest at the time of analysis, with approximately 38.42% censored observations. A cut-off point of 42 months was applied to define the observation period. The event variable was defined as graduation status, where a value of 1 indicates that a student had graduated and 0 indicates that the student had not yet graduated or censored. The survival time variable represents the duration of study, measured continuously in months until graduation or censoring.

The independent variables consist of demographic and academic characteristics that may influence study duration, including gender, cohort year, faculty, region of origin, type of undergraduate institution, accreditation level of undergraduate institution, age, undergraduate GPA, and marital status. Several feature engineering procedures were additionally applied to improve the representation of nonlinear relationships and interactions within the data, including quadratic and interaction-based features. Categorical variables were encoded prior to model development to ensure compatibility with the implemented machine learning models.

This study utilizes four tree-based survival models, namely ST, EST, RSF, and GBS. These models were selected because of their capability to model nonlinear relationships and complex interactions without relying on strict parametric assumptions. Traditional statistical approaches such as the Cox proportional hazards model were not included because the proportional hazards assumption may not adequately represent heterogeneous educational datasets. Model performance was evaluated using the C-index and IBS to assess discrimination and prediction accuracy over time.

The research procedure began with data preprocessing, including data cleaning, handling missing values, categorical encoding, and feature engineering. Feature engineering included quadratic transformation of age and interaction-based variables to better represent potential nonlinear effects and complex relationships among predictors. The dataset was then divided into training and testing sets with proportion 80% training data and 20% testing data using a repeated stratified random split approach with 10 repetitions while preserving the proportion of censored and uncensored observations in each split. Stratification was performed based on the event indicator to preserve the proportion of graduated and censored observations across the training and testing datasets in each repetition. Subsequently, each survival model was trained using the training dataset and evaluated using the testing dataset based on C-index and IBS metrics. The best-performing model was further utilized for risk stratification analysis by grouping students into low-, medium-, and high-risk categories based on predicted survival probabilities at a 24-month time horizon using a quantile-based grouping approach.

RESULT AND DISCUSSION

This section presents the comparative performance of the evaluated tree-based survival models, followed by student risk stratification analysis and a discussion of the findings. The results are organized to demonstrate both the predictive capability of the proposed models and their practical implications for academic monitoring in higher education.

Model performance comparison

Table 1 summarizes the predictive performance of the four tree-based survival models evaluated using repeated stratified random splits with 10 repetitions. The results indicate performance differences across models and evaluation metrics. Among all evaluated methods, GBS achieved the highest mean C-index (0.659 ± 0.014) and the lowest mean IBS (0.194 ± 0.005), indicating the best overall predictive performance. RSF showed competitive discrimination performance with a slightly lower mean C-index, while ST and EST produced lower overall predictive accuracy.

Table 1. Model performance comparison

Model	Mean C-index ± SD	Mean IBS ± SD
ST	0.633 ± 0.014	0.198 ± 0.005
EST	0.622 ± 0.021	0.208 ± 0.008
RSF	0.658 ± 0.016	0.211 ± 0.005
GBS	0.659 ± 0.014	0.194 ± 0.005

As presented in Table 1, GBS achieved the highest predictive discrimination and the lowest prediction error among the evaluated models, whereas ST, EST, and RSF exhibited different trade-offs between discrimination and calibration performance. The comparison results demonstrate different predictive characteristics among tree-based survival models. GBS consistently produced strong performance across both evaluation metrics, suggesting that boosting-based approaches are more effective in capturing complex nonlinear relationships and interactions within postgraduate educational data. RSF also achieved high discrimination performance but showed weaker calibration compared to GBS. In contrast, the ST model demonstrated relatively stable performance across repeated evaluations despite lower predictive capability, while EST exhibited greater variability and sensitivity to data partitioning.

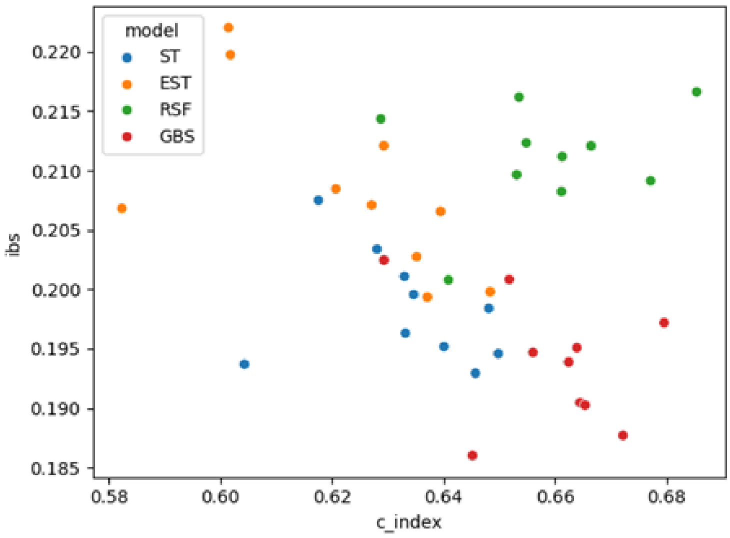

Figure 2. Scatter plot of mean c-index and IBS across models

As illustrated in Figure 2, GBS achieved the most favorable balance between predictive discrimination and calibration, while the remaining models showed different trade-offs between the two evaluation metrics. Figure 3 presents the distribution of model performance across repeated evaluations using boxplots. The repeated stratified random split approach provides a more robust evaluation framework by reducing dependence on a single train–test partition and maintaining the proportion of censored observations in each repetition. The boxplots show that ensemble-based models generally achieved better predictive performance than single-tree approaches, although differences in variability remained across models. Overall, the repeated evaluation results indicate that GBS provided the most balanced and consistent predictive performance among the evaluated models.

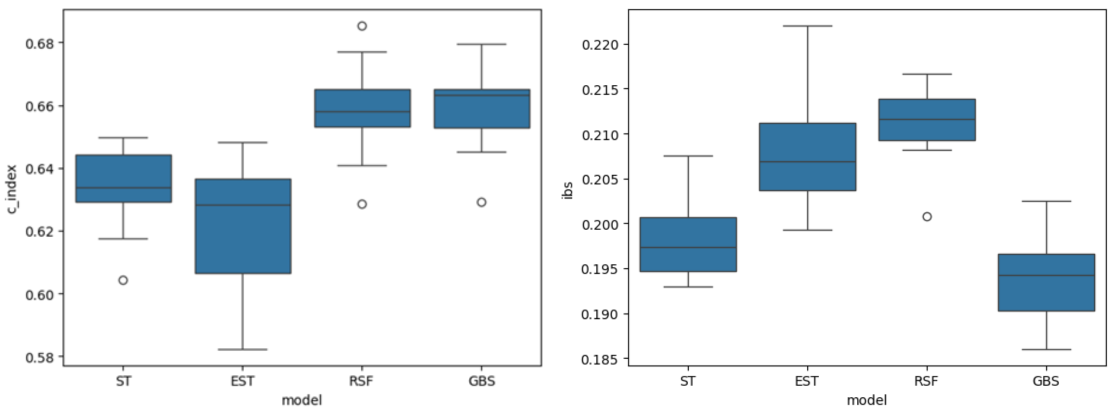

Figure 3. Boxplot of model performance across repeated evaluations

Risk Stratification Analysis

Based on the best-performing model, GBS, risk stratification was conducted using predicted survival probabilities at a predefined 24-month time horizon. Students were grouped into low, medium, and high-risk categories using a quantile-based approach to ensure balanced group distribution while maintaining meaningful separation in risk levels. Figure 4 shows the Kaplan–Meier survival curves for the three risk groups. The curves demonstrate clear separation patterns, indicating that the stratification effectively distinguishes students with different probabilities of delayed graduation. The log-rank test further confirmed significant differences in survival distributions among the risk groups.

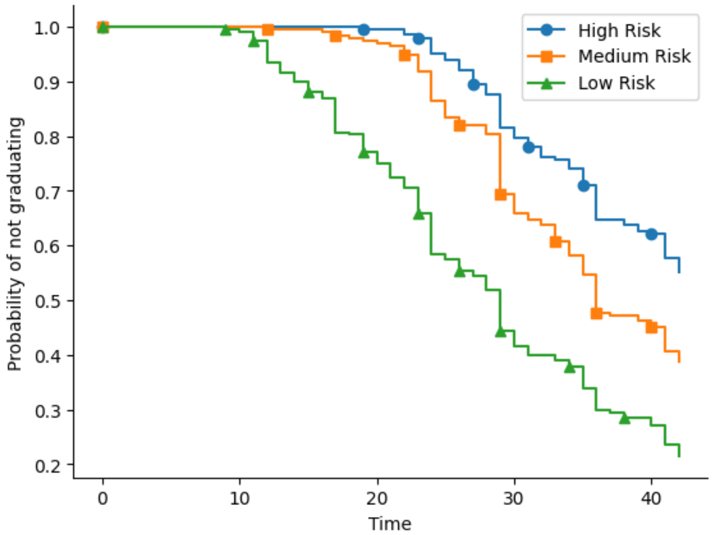

Figure 4. Risk stratification at 24 months

The stratification results reveal several meaningful patterns associated with study duration. Table 2 shows that students in the high-risk group tended to have higher average age and lower undergraduate GPA compared to students in the low-risk group. These findings indicate that academic readiness and demographic characteristics remain important factors associated with delayed graduation. In addition, the distribution of risk groups varied across demographic and institutional backgrounds. From Figure 5, we know that male students and students from private undergraduate institutions appeared more frequently in the high-risk category, while female students and graduates from public institutions were more dominant in the low-risk group. These patterns suggest that the model successfully captures heterogeneous characteristics within postgraduate educational data.

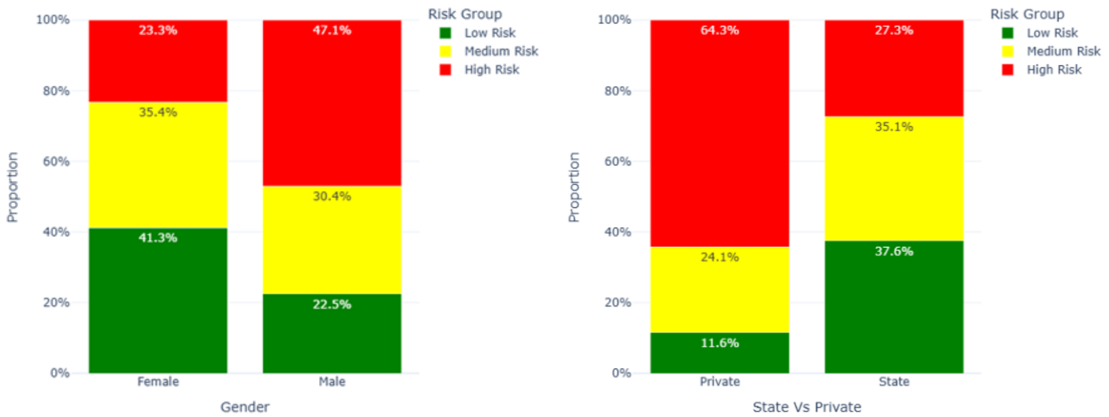

Figure 5. Distribution of risk groups by gender and undergraduate institution type

Table 2. Characteristics of risk groups

Risk Group	Mean Age	Mean Undergraduate GPA
Low Risk	22.85	3.62
Medium Risk	25.99	3.45
High Risk	29.01	3.12

Overall, the integration of survival prediction and risk stratification enhances the interpretability and practical relevance of the proposed framework. The ability of the GBS model to separate students into distinct risk categories demonstrates that tree-based survival machine learning methods can provide not only predictive performance but also meaningful student profiling in educational time-to-event analysis.

Discussion and implications

The findings demonstrate the growing relevance of machine learning approaches in survival analysis and educational predictive analytics. Tree-based survival models provide flexible frameworks for modeling complex time-to-event data without requiring strict parametric assumptions [20]. Previous studies in educational data mining and predictive analytics have also shown that machine learning methods substantially improve the prediction of student outcomes and academic performance [6], [21]. In this study, ensemble-based survival models consistently outperformed single-tree approaches, indicating the importance of handling nonlinear relationships and heterogeneous educational characteristics within postgraduate student data.

The observed performance differences are also closely related to the characteristics of the dataset. The educational dataset used in this study consists of heterogeneous demographic and academic variables, including both numerical and categorical predictors, with potentially nonlinear relationships to study duration. Such characteristics are well suited to ensemble tree-based survival methods because they can automatically capture complex interactions without relying on proportional hazards or linearity assumptions. Consequently, GBS demonstrated superior predictive performance by iteratively reducing prediction errors while effectively modeling the heterogeneous structure of the educational data.

Among the evaluated models, GBS achieved the best overall predictive performance, demonstrating its effectiveness in capturing subtle patterns associated with study duration. The boosting mechanism iteratively minimizes prediction errors, enabling better discrimination and calibration performance across repeated evaluations [22]. In contrast, RSF also demonstrated strong predictive capability through bootstrap aggregation and ensemble learning, which improve model robustness and reduce variance, However, its higher IBS suggests that strong discrimination performance does not necessarily guarantee optimal calibration over time [23], [24]. Comparative studies across survival modeling applications similarly report that ensemble-based approaches generally outperform simpler tree-based methods in complex prediction tasks [25]. In addition, recent developments in scalable and hybrid survival machine learning frameworks further highlight the growing capability of advanced predictive models for structured survival data [26], [27].

The repeated stratified random split evaluation used in this study also emphasizes the importance of robustness assessment in survival machine learning research. Although RSF achieved competitive discrimination performance, GBS provided a better balance between discrimination and calibration across repeated experiments. These findings suggest that model evaluation should not rely solely on a single performance metric or a single train–test partition. The observed variability among models further confirms that predictive performance in survival analysis is highly dependent on dataset characteristics and model learning strategies [28]. Compared with conventional approaches, survival machine learning frameworks provide more adaptive and data-driven solutions for modeling longitudinal educational outcomes.

The integration of survival prediction and risk stratification also enhances the practical relevance of the proposed framework in higher education settings. The ability to identify students with different levels of graduation risk enables institutions to implement more targeted academic monitoring and intervention strategies. Previous studies in learning analytics and educational data science similarly emphasize the growing role of predictive systems in supporting student success, academic planning, and institutional decision-making [29], [7], [8]. Broader applications of machine learning and educational data mining further demonstrate the increasing adoption of data-driven approaches in educational systems [30]. In this context, the proposed survival machine learning framework offers a more proactive approach for identifying students who may require additional academic support during their study period. From a practical perspective, the proposed framework is most beneficial when applied during the early to middle stages of graduate study, before students approach the expected graduation period. Early prediction enables universities to identify students with an elevated risk of delayed graduation while sufficient time remains for academic advising, mentoring, or other targeted interventions. Consequently, the framework may support proactive rather than reactive academic monitoring. Accordingly, the proposed framework is particularly suitable for educational datasets containing heterogeneous student characteristics, mixed numerical and categorical predictors, and right-censored time-to-event outcomes.

This study has several limitations. The analysis was conducted using data from a single institution, which may limit the generalizability of the findings to other educational environments. In addition, the study only considered structured demographic and academic variables, while other potentially relevant factors such as behavioral, financial, and psychological characteristics were not included. External validation using independent educational datasets was also not conducted in this study. Future studies may explore multicenter educational datasets, additional predictors, and more advanced survival modeling approaches to further improve predictive performance and model generalizability.

CONCLUSION

This study demonstrates that tree-based survival machine learning provides an effective framework for predicting graduate study duration while supporting meaningful student risk stratification in higher education. Among the evaluated models, Gradient Boosting Survival showed the most reliable overall predictive performance, indicating the potential of ensemble-based survival methods to model complex educational time-to-event data. Beyond model comparison, the integration of survival prediction with risk stratification offers a practical approach for identifying students who may require earlier academic support, thereby enabling more proactive and evidence-based academic monitoring. These findings contribute to the growing application of survival machine learning in educational analytics by providing both methodological guidance for model selection and practical insights for institutional decision-making. Future studies may further improve the generalizability of this framework through external validation using multicenter educational datasets and additional behavioral or socioeconomic predictors.

ACKNOWLEDGMENTS

The authors thank IPB University for providing access to the data used in this study. ChatGPT (OpenAI) was used to assist with English language editing and improving the clarity of the manuscript. All scientific content, analyses, interpretations, and conclusions were developed and verified by the authors, who take full responsibility for the final manuscript.

REFERENCES

[1] D. A. Shafiq, M. Marjani, R. A. A. Habeeb, and D. Asirvatham, "Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review," *IEEE Access*, vol. 10, pp. 72480–72503, 2022, doi: 10.1109/ACCESS.2022.3188767.

[2] A. Boroowa and C. Herodotou, "Learning Analytics in Open and Distance Higher Education: The Case of the Open University UK," in *Learning Analytics in Open and Distributed Learning: Potential and Challenges*, P. Prinsloo, S. Slade, and M. Khalil, Eds., Singapore: Springer Nature Singapore, 2022, pp. 47–62. doi: 10.1007/978-981-19-0786-9_4.

[3] D. Zhang, "Research on Dance Art Teaching System Based on Data Mining and Machine Learning," *Comput. Aided. Des. Appl.*, pp. 54–68, Jul. 2023, doi: 10.14733/cadaps.2024.S2.54-68.

[4] A. Stojanov and B. K. Daniel, "A decade of research into the application of big data and analytics in higher education: A systematic review of the literature," *Educ. Inf. Technol. (Dordr).*, vol. 29, no. 5, pp. 5807–5831, 2024, doi: 10.1007/s10639-023-12033-8.

[5] Z. Ersozlu, S. Taheri, and I. Koch, "A review of machine learning methods used for educational data," *Educ. Inf. Technol. (Dordr).*, vol. 29, no. 16, pp. 22125–22145, 2024, doi: 10.1007/s10639-024-12704-0.

[6] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H.-Y. Kwon, and A. Hussain, "Educational data mining to predict students' academic performance: A survey study," *Educ. Inf. Technol. (Dordr).*, vol. 28, no. 1, pp. 905–971, Jan. 2023, doi: 10.1007/s10639-022-11152-y.

[7] K. Fahd and S. J. Miah, "Designing and evaluating a big data analytics approach for predicting students' success factors," *J. Big Data*, vol. 10, no. 1, p. 159, Oct. 2023, doi: 10.1186/s40537-023-00835-z.

[8] H. T. Adu-Twum, E. A. Sarfo, and E. Nartey, "Role of Advanced Data Analytics in Higher Education: Using Machine Learning Models to Predict Student Success," *Journal of Data Science and Artificial Intelligence*, vol. 3, no. 1, 2024.

[9] J. Luo, L. Xie, H. Yang, X. Yin, and Y. Zhang, "Machine Learning for Time-to-Event Prediction and Survival Clustering: A Review from Statistics to Deep Neural Networks," in *Communications*

in Computer and Information Science: Intelligent Computers, Algorithms, and Applications, Singapore: Springer Nature, 2024, pp. 174–192. doi: 10.1007/978-981-97-0065-3_13.

[10] Z.-H. Peng, Z.-X. Huang, J.-H. Tian, T. Chong, and Z.-L. Li, "The application of time-to-event analysis in machine learning prognostic models," *J. Transl. Med.*, vol. 22, no. 1, p. 146, Feb. 2024, doi: 10.1186/s12967-024-04909-1.

[11] S. Wiegrebe, P. Kopper, R. Sonabend, B. Bischl, and A. Bender, "Deep learning for survival analysis: a review," *Artif. Intell. Rev.*, vol. 57, no. 3, p. 65, Feb. 2024, doi: 10.1007/s10462-023-10681-3.

[12] 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.

[13] K. Evangeline I., S. P. A. Kirubha, and J. G. Precious, "Survival analysis of breast cancer patients using machine learning models," *Multimed. Tools Appl.*, vol. 82, no. 20, pp. 30909–30928, Aug. 2023, doi: 10.1007/s11042-023-14989-8.

[14] J. Lin *et al.*, "The Development of a Prediction Model Based on Random Survival Forest for the Postoperative Prognosis of Pancreatic Cancer: A SEER-Based Study," *Cancers (Basel).*, vol. 14, no. 19, p. 4667, Sep. 2022, doi: 10.3390/cancers14194667.

[15] M. Bai, Y. Zheng, and Y. Shen, "Gradient boosting survival tree with applications in credit scoring," *Journal of the Operational Research Society*, vol. 73, no. 1, pp. 39–55, Jan. 2022, doi: 10.1080/01605682.2021.1919035.

[16] L. Burk, J. Zobolas, B. Bischl, A. Bender, M. N. Wright, and R. Sonabend, "A large-scale neutral comparison study of survival models on low-dimensional data," *Bioinformatics*, vol. 42, no. 5, May 2026, doi: 10.1093/bioinformatics/btag186.

[17] M. S. Zaenal, A. Fitrianto, and H. Wijayanto, "Comparison of Extremely Randomized Survival Trees and Random Survival Forests: A Simulation Study," *Scientific Journal of Informatics*, vol. 11, no. 3, pp. 635–644, Aug. 2024, doi: 10.15294/sji.v11i3.8464.

[18] W. Lin, D. Shu, N. Geng, D. Zhang, C. Feng, and S. Chen, "DeepSurv-based deep learning model for survival prediction and personalized treatment recommendation in tongue squamous cell carcinoma," *Journal of Cranio-Maxillofacial Surgery*, vol. 53, no. 9, pp. 1334–1343, Sep. 2025, doi: 10.1016/j.jcms.2025.05.005.

[19] T. Chen, "Survival Analysis Revisited: Understanding and Unifying Poisson, Exponential, and Cox Models in Fall Risk Analysis," Jan. 2025.

[20] S. Salerno and Y. Li, "High-Dimensional Survival Analysis: Methods and Applications," *Annu. Rev. Stat. Appl.*, vol. 10, no. 1, pp. 25–49, Mar. 2023, doi: 10.1146/annurev-statistics-032921-022127.

[21] E. Ahmed, "Student Performance Prediction Using Machine Learning Algorithms," *Applied Computational Intelligence and Soft Computing*, vol. 2024, no. 1, Jan. 2024, doi: 10.1155/2024/4067721.

[22] A. Archetti, E. Lomurno, D. Piccinotti, and M. Matteucci, "FPBoost: Fully Parametric Gradient Boosting for Survival Analysis," Feb. 2026, doi: 10.3233/FAIA251019.

[23] R. Zhu, S. E. Formentini, and Y. Cui, "Random Forests for Survival Analysis and High-Dimensional Data," 2023, pp. 831–847. doi: 10.1007/978-1-4471-7503-2_40.

[24] L. Zhang *et al.*, "Prediction of prognosis in elderly patients with sepsis based on machine learning (random survival forest)," *BMC Emerg. Med.*, vol. 22, no. 1, p. 26, Dec. 2022, doi: 10.1186/s12873-022-00582-z.

[25] S. Mishra, "A Comparative Study for Time-to-Event Analysis and Survival Prediction for Heart Failure Condition using Machine Learning Techniques," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 4, no. 3, pp. 115–134, Jul. 2022, doi: 10.35882/jeeemi.v4i3.225.

[26] S. Babaei Rikan, A. Sorayaie Azar, A. Naemi, J. Bagherzadeh Mohasefi, H. Pirnejad, and U. K. Wiil, "Survival prediction of glioblastoma patients using modern deep learning and machine learning techniques," *Sci. Rep.*, vol. 14, no. 1, p. 2371, Jan. 2024, doi: 10.1038/s41598-024-53006-2.

[27] K. P, S. P. S, K. N, and V. P. Balakrishnan, "Utilizing Cloud Computing to Deliver Accurate and Scalable Diabetes Predictions with XGBoost," in *2024 Second International Conference on Advances in Information Technology (ICAIT)*, IEEE, Jul. 2024, pp. 1–7. doi: 10.1109/ICAIT61638.2024.10690419.

[28] W. Tizi and A. Berrado, “Machine learning for survival analysis in cancer research: A comparative study,” *Sci. Afr.*, vol. 21, p. e01880, Sep. 2023, doi: 10.1016/j.sciaf.2023.e01880.

[29] O. Viberg, L. Engström, M. Saqr, and S. Hrastinski, “Exploring students’ expectations of learning analytics: A person-centered approach,” *Educ. Inf. Technol. (Dordr).*, vol. 27, no. 6, pp. 8561–8581, 2022, doi: 10.1007/s10639-022-10980-2.

[30] C. Baek and T. Doleck, “Educational Data Mining versus Learning Analytics: A Review of Publications From 2015 to 2019,” *Interactive Learning Environments*, vol. 31, no. 6, pp. 3828–3850, Aug. 2023, doi: 10.1080/10494820.2021.1943689.