JAIEP

Predicting Employee Retention Using Artificial Intelligence and Survival Analysis Approaches

Serifat A Folorunso[1]*, Richard O. Kehinde[2], Morufu A. Folorunso[3]

[1] University of Ibadan, Laboratory for Statistical Analysis (UI-LISA), Department of Statistics, Nigeria

[2] Federal College of Animal Health and Production Technology, Department of Statistics, Moor Plantation, Ibadan, Oyo State, Nigeria

[3] Federal School of Statistics, General Studies Department, Sasa Road, Ibadan, Nigeria

* Corresponding author: serifatf005@gmail.com

Abstract

Background: Employee retention is a critical concern for organizations seeking to maintain a stable and productive workforce. Understanding the factors driving turnover is essential for designing effective HR interventions. Methods: This study applies advanced survival analysis techniques, including the Kaplan–Meier estimator, Cox proportional hazards model, and Random Survival Forests (RSF), to predict employee retention and identify key determinants of turnover. Data from 1,480 employees, with 16% (238) having left the organization, were analyzed, considering variables such as age, job satisfaction, overtime, and departmental affiliation. Results: Employees working overtime are 3.57 times more likely to leave, indicating overtime as a major risk factor. In contrast, older employees and those with higher job satisfaction show reduced turnover risks, with hazard ratios of 0.93 and 0.79, respectively. Departmental differences were observed, with Research & Development exhibiting the highest retention, including employees staying beyond 30 years, while Human Resources had the highest turnover, particularly within the first five years. Conclusion: The study highlights the importance of job satisfaction, overtime policies, and department-specific retention strategies in reducing turnover. Survival modeling provides actionable insights into the timing and drivers of employee exits, enabling organizations to implement targeted interventions that enhance workforce stability and long-term organizational success.

Keywords: Kaplan–Meier, Predictive Analytics, hazards model; Random survival forest; Workforce sustainability.

MSC: 68M25; 60G35

Doi : https://doi.org/10.21608/jaiep.2025.431657.1030

Received: May 22, 2025, Revised: June 29, 2025, Accepted: July 8, 2025

1. Introduction

The rapid expansion of next-generation sequencing technologies and the associated drop in genome sequencing costs have led to an unprecedented influx of biological data [1]. This has transformed DNA sequence analysis into a central component of biomedical research and personalized healthcare. Fundamental to many bioinformatics tasks is the ability to identify subsequences — or patterns — within vast genomic datasets. This problem has direct applications in mutation detection, disease classification, forensic analysis, and evolutionary inference.

Employee retention significantly influences organizational performance, innovation, and financial health. High turnover rates increase recruitment and training costs, diminish employee morale, and disrupt workflow (Alam & Asim, [1]; An [4]; Lee [8]; Chatwick, [5]). Traditional analytical approaches often fail to incorporate the timing of employee exits, providing limited guidance for proactive interventions (Kurdi & Alshurideh, [7]). This study combines AI techniques and survival analysis to model retention patterns and predict when employees are likely to leave. The integration of Kaplan-Meier estimation, Cox Proportional Hazards (Cox-PH) modeling, and Random Survival Forests (RSF) offers both interpretability and predictive accuracy for employee turnover analysis.

2. Related Work

Previous studies identify age, wage, and satisfaction as major turnover determinants (Madariaga et al., [9]; Rahimi et al., [11]). AI models, including decision trees and random forests, have improved attrition prediction accuracy (El-Rayes et al., [6]), but often neglect time-to-event censoring. Survival analysis addresses this by modeling both event occurrence and duration (Nasejje et al., [10]). This study extends the literature by fusing AI's predictive power with survival analysis's temporal insight, offering a richer interpretation of employee lifecycle patterns, see Alexander [2], and Almerri [3].

3. Methodology

The methodology section outlines the dataset, analytical framework, and tools used in this study.

3.1 Data Description

The dataset comprises 1,480 employees with 38 variables, including demographics, compensation, and satisfaction measures. Key predictors include Age, Gender, Monthly Income, Job Satisfaction, Work-Life Balance, Years Since Last Promotion, Distance from Home, Number of Companies Worked, Job Level, and Overtime. Dependent variables: Survival Time (tenure until exit or censoring) and Event (1 = left, 0 = retained).

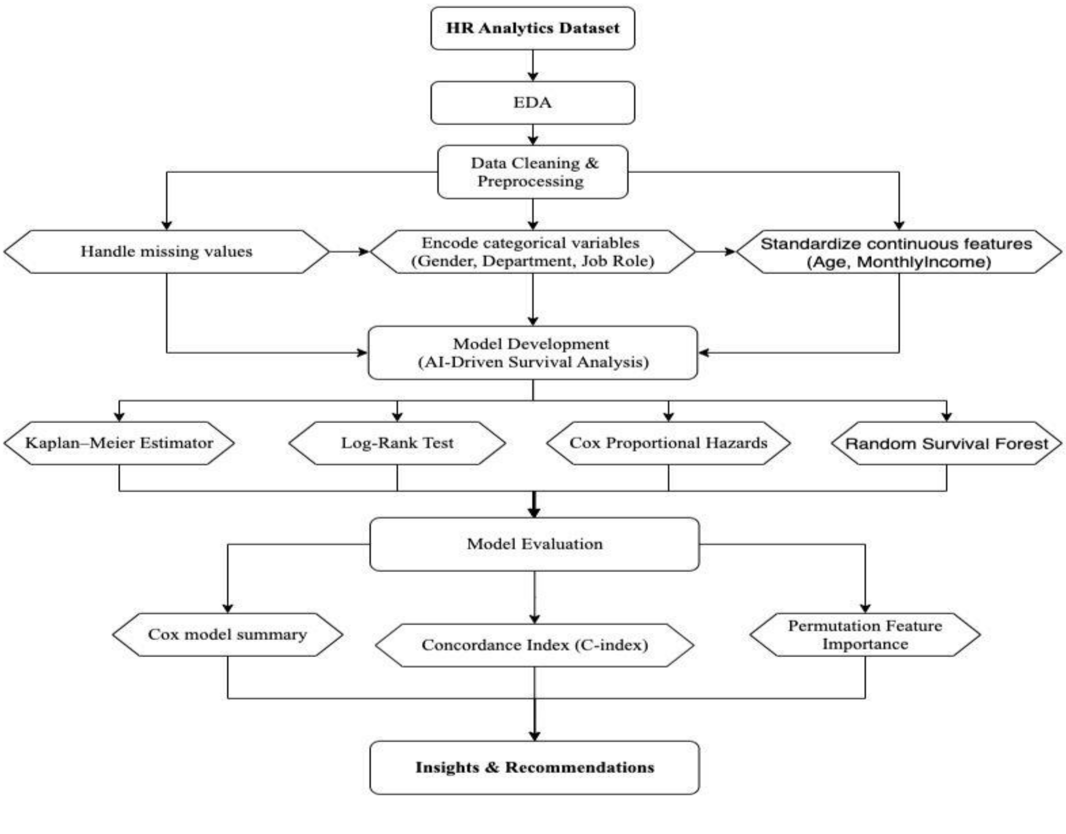

Figure 1. Analytical Framework of the Study: This diagram illustrates the integration of AI and survival analysis techniques.

3.2 Analytical Framework

Four analytical models were employed: Kaplan-Meier Estimator, Log-Rank Test, Cox Proportional Hazards Model, and Random Survival Forest (RSF). Figure 1 above describes in detail the Flowchart of Analytical Procedure, which showcases the Workflow details such as data cleaning, feature selection, KM estimation, Cox regression, and RSF modeling.

3.3 Software Tools

Analysis was conducted in Python using lifelines and scikit-survival. Model performance was evaluated through the concordance index (C = 0.83), AIC, and log-likelihood tests.

3.4 Overview of Analysis

3.4.1 Exploratory Data Analysis

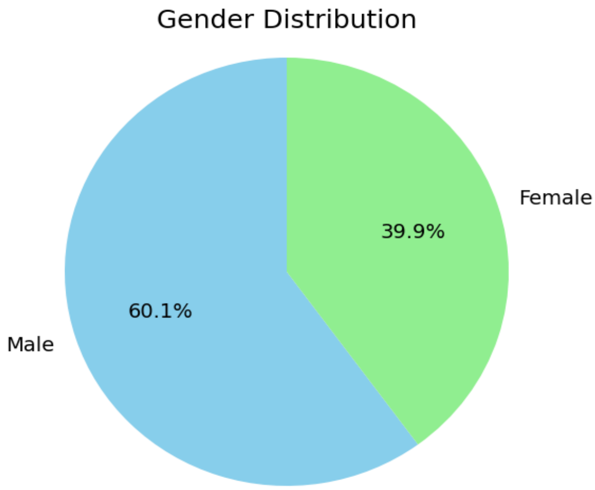

Figure 2: Distribution of Gender

The pie chart above shows the gender distribution of a group where males make up 60.1% of the group, while females constitute 39.9%. The chart visually represents this breakdown, with a larger portion of the pie representing males and a smaller portion representing females.

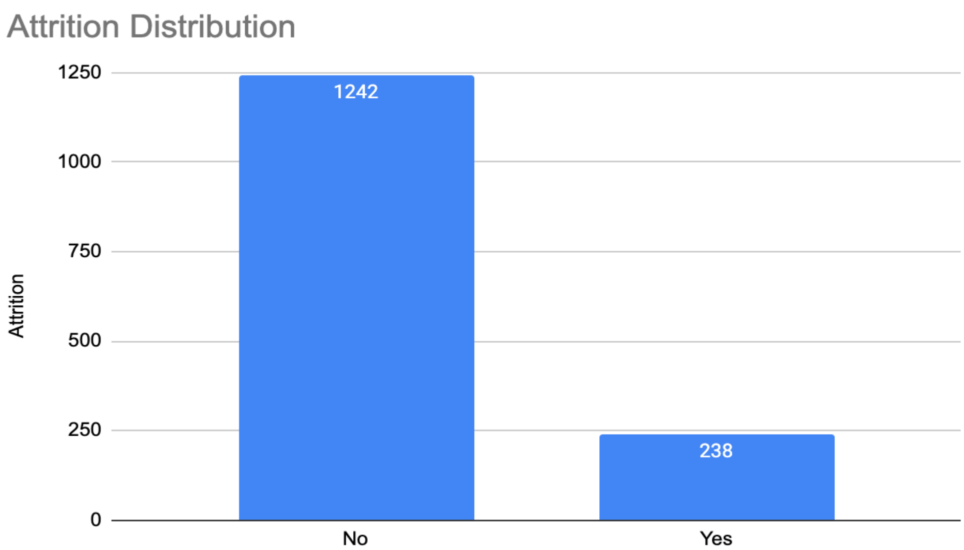

Figure 3: Distribution of Attrition

Doi : https://doi.org/10.21608/jaiep.2025.431657.1030
May 22, 2025; Revised: June 29, 2025; Accepted: July 8, 2025

In this dataset, employee attrition indicates that out of a total of 1,480 employees, 238 have left the company, accounting for approximately 16%, while 1,242 (84%) have remained. This highlights a moderate employee turnover rate.

3.5 Kaplan-Meier Estimator for Survival Function

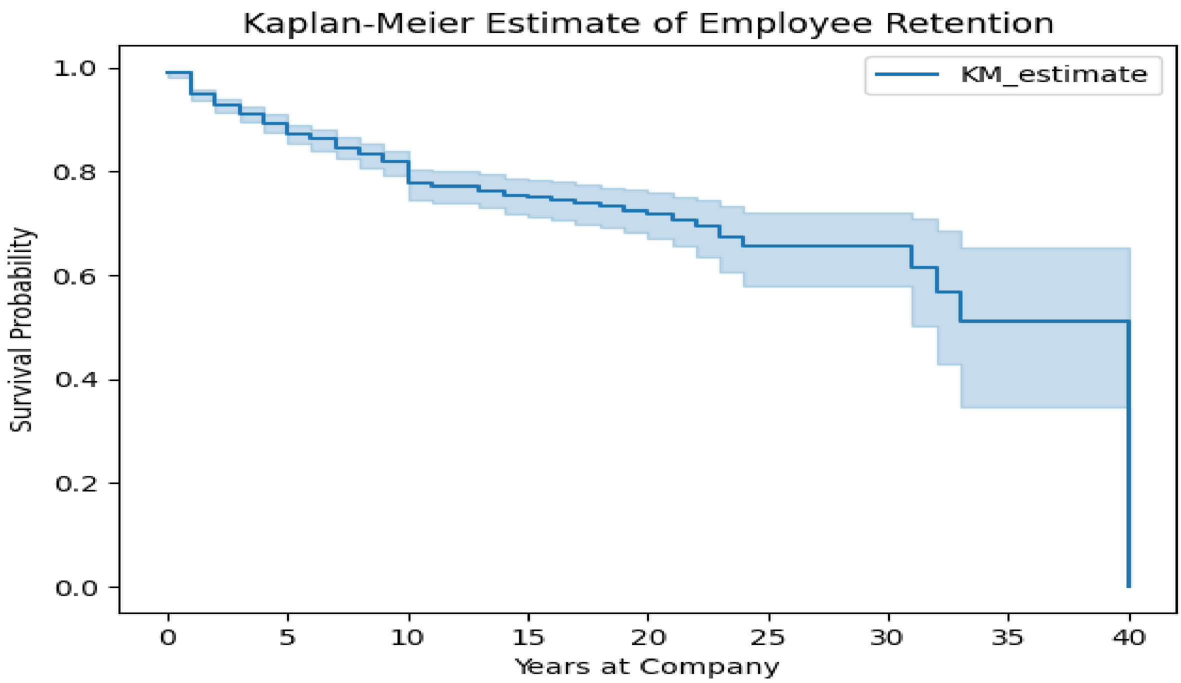

Figure 4: Survival analysis of employee retention

The Kaplan-Meier curve illustrates the probability of an employee remaining with the company over time. The survival probability starts at 1, indicating that all employees are initially employed. As the years at the company increase, the survival probability gradually decreases, suggesting that the likelihood of an employee staying decreases with tenure. The curve shows a plateau around 30 years, indicating that a significant portion of employees who reach this milestone tend to stay for a prolonged period. However, beyond 35 years, the survival probability drops sharply, suggesting a higher turnover rate among long-tenured employees.

3.6 Develop a Survival Model (Cox Proportional Hazards Model)

The following tables 1 and 2 present the results of the Cox Proportional Hazards regression analysis on employee turnover and its determinants. The coefficients indicate both the direction and magnitude of each factor's influence on the likelihood of turnover—negative values imply a reduced risk of leaving, while positive values indicate an increased risk.

Overall, Age, Job Satisfaction, and Years Since Last Promotion are negatively associated with turnover, suggesting that older, more satisfied employees and those recently promoted are less likely to leave. Monthly Income shows no meaningful effect on retention, indicating that salary alone does not predict turnover. Distance From Home has a slight positive association with retention, while Work-Life Balance and Environment Satisfaction also show modest negative relationships with turnover. Notably, employees working overtime (Overtime Yes) display a significantly higher turnover risk, with the largest coefficient (1.27) and hazard ratio (3.572), highlighting overtime as a key driver of employee exit.

Table 1: Cox Proportional Hazards Model Results for Employee Turnover

	coef	exp(coef)	se(coef)	z	p	-log2(p)
Age	-0.07	0.93	0.01	-6.47	<0.005	33.21
MonthlyIncome	0	1	0	-0.89	0.37	1.43
JobSatisfaction	-0.24	0.79	0.06	-4.1	<0.005	14.58
DistanceFromHome	0.03	1.03	0.01	3.67	<0.005	11.99
YearsSinceLastPromotion	-0.09	0.92	0.03	-3.37	<0.005	10.35
WorkLifeBalance	-0.19	0.82	0.09	-2.14	0.03	4.96
NumCompaniesWorked	0.17	1.19	0.02	6.96	<0.005	38.07
JobLevel	-0.63	0.53	0.2	-3.09	<0.005	8.99
EnvironmentSatisfaction	-0.22	0.8	0.06	-3.81	<0.005	12.79
OverTime_Yes	1.27	3.57	0.13	9.65	<0.005	70.77

This table 1 presents key performance metrics assessing the overall quality and reliability of the Cox Proportional Hazards model. The concordance index of 0.83 demonstrates strong predictive accuracy, indicating the model effectively distinguishes between employees likely to stay and those likely to leave. The partial AIC value of 2863.13 reflects a good model fit, with lower scores representing better performance. Additionally, the log-likelihood ratio test statistic of 334.42 (on 16 degrees of freedom) and the -log2(p) value of 201.78 confirm the statistical significance and robustness of the model in explaining employee turnover patterns.

Table 2: Model Evaluation Metrics for the Cox Proportional Hazards Model

Concordance	Partial AIC	log-likelihood ratio test	-log2(p) of ll-ratio test
0.83	2863.13	334.42 on 16 df	201.78

3.7 Random Survival Forest

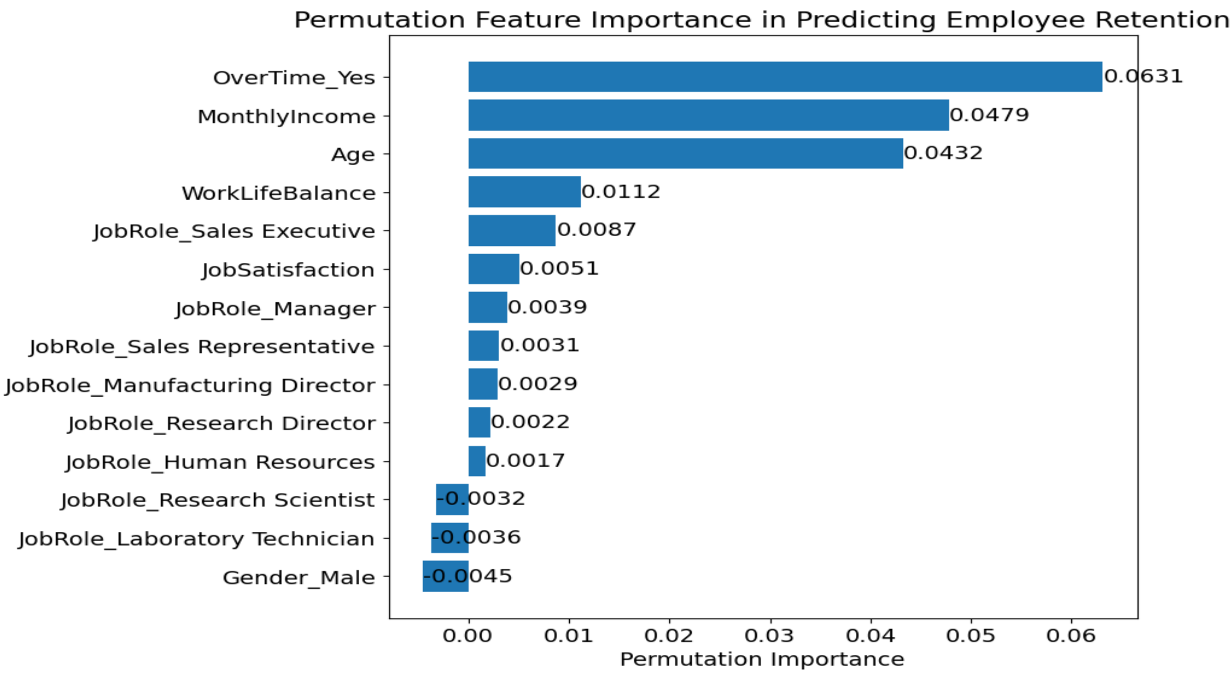

Figure 5: Random Survival Forest Permutation Feature Importance Plot

Doi : https://doi.org/10.21608/jaiep.2025.431657.1030
May 22, 2025; Revised: June 29, 2025; Accepted: July 8, 2025

The permutation feature importance results from the Random Survival Forest (RSF) model show that OverTime_Yes is the most influential predictor of employee turnover, with an importance score of 0.0631, followed by Monthly Income (0.0479) and Age (0.0432). In contrast, Work-Life Balance (0.0112) and Job Satisfaction (0.0051) have relatively minor effects, while job roles such as Sales Executive and Manager show minimal influence, with scores below 0.01. Negative importance values for roles like Research Scientist and Laboratory Technician, as well as for Gender_Male, suggest potential links to higher turnover risk, highlighting the complex interplay of factors influencing employee retention.

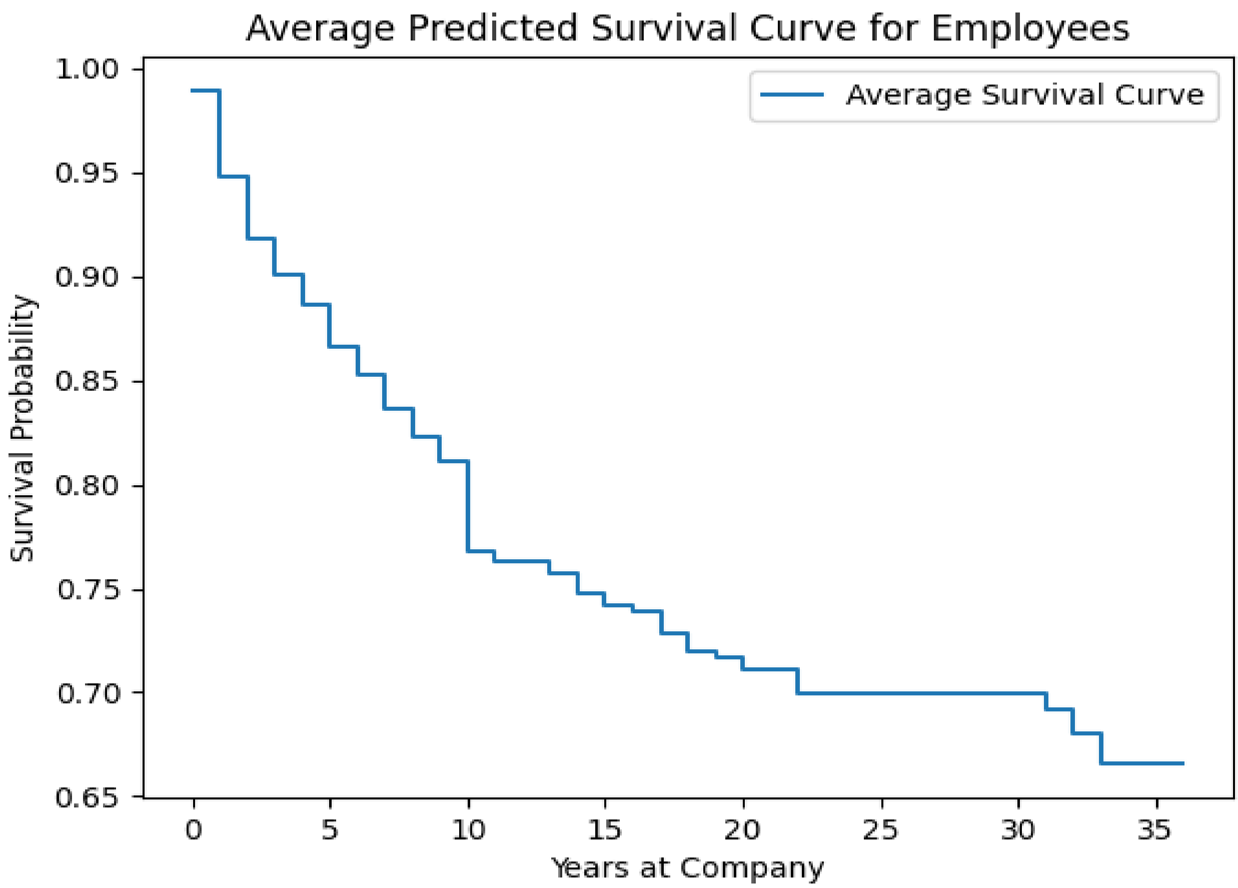

Figure 6: Average Predicted Survival Curve for Employees based on the Random Survival Forest Model

The average predicted survival curve from the Random Survival Forest model depicts the probability of employees remaining with the company over time. Initially, the survival probability is 1.0, representing full employment. As tenure increases, the probability gradually declines, indicating a steady rise in turnover risk. The curve plateaus around 25 years, suggesting that employees who reach this point are likely to remain long-term. However, a sharp decline beyond 30 years reflects increased turnover among long-tenured staff. Overall, the curve captures key retention patterns, illustrating typical employee tenure and highlighting areas where retention strategies could be strengthened.

4. Summary of Analytical Findings

The exploratory data analysis (EDA) reveals key demographic and behavioral patterns underlying employee turnover. The workforce comprises 60.1% males and 39.9% females, with a mean age of 36.9 years and a moderate turnover rate of 16% (238 of 1,480 employees). Most employees

(71.76%) do not work overtime, and the majority are concentrated in the Research & Development department (65.34%), followed by Sales (30.41%) and Human Resources (4.26%).

The Kaplan-Meier survival curves indicate a gradual decline in retention with increasing tenure, stabilizing around 30 years before a marked drop beyond 35 years. Turnover is higher among employees working overtime and among women compared to men. Departmental analysis shows the highest retention in Research & Development and the lowest in Human Resources, while job role analysis suggests greater stability among Laboratory Technicians and Sales Representatives compared to Managers and Manufacturing Directors.

The Cox Proportional Hazards model identifies Age (-0.07), Job Satisfaction (-0.24), and Years Since Last Promotion (-0.09) as factors negatively associated with turnover risk, whereas OverTime_Yes (1.27) significantly increases the likelihood of leaving. Model diagnostics confirm strong predictive accuracy (concordance = 0.83) and good fit ($\chi^2 = 334.42$, $p < 0.001$).

Findings from the Random Survival Forest (RSF) further emphasize OverTime_Yes (0.0631), Monthly Income (0.0479), and Age (0.0432) as the most influential predictors of turnover, with Work-Life Balance (0.0112) and Job Satisfaction (0.0051) showing lower importance. The RSF survival curve aligns with Kaplan-Meier results, depicting declining retention with tenure and a sharp decrease after 30 years.

Overall, the analyses underscore that overtime, income, age, job satisfaction, and promotion opportunities are key determinants of turnover. Addressing these factors through improved work-life balance policies, career progression structures, and overtime management could substantially enhance employee retention and organizational stability.

5. Discussion and Conclusion

This study examined the factors influencing employee turnover and retention, focusing on demographic characteristics, work environment, and job-related attributes. It also explored the optimal timing for HR interventions aimed at reducing turnover. The findings reveal significant associations between variables such as age, job status, income, job satisfaction, and working conditions with turnover rates across organizations. Key determinants of retention include job satisfaction, organizational culture, career advancement opportunities, and workload, all of which play vital roles in employees' decisions to remain or leave.

By applying survival analysis models and other quantitative techniques, the study effectively identified both the likelihood and timing of employee turnover events. Comparative analysis of retention strategies further highlighted the importance of targeted interventions, particularly in high-risk departments and job categories. Enhancing job satisfaction, providing clear career pathways, and fostering a positive organizational climate were found to significantly improve retention outcomes. A stable workforce not only enhances productivity but also promotes morale and organizational resilience, creating long-term benefits for both employers and employees.

While the study offers valuable insights, it also emphasizes the need to consider industry-specific contexts and advanced predictive methods in future research. These areas remain crucial for refining HR practices and developing effective, evidence-based retention strategies across diverse organizational settings.

5.2 Recommendations for Future Research

Future studies should investigate how industry characteristics influence turnover and retention dynamics. Comparative analyses across organizational levels and job types could further clarify which HR interventions are most effective in reducing turnover. Employing larger and more diverse samples, as well as longitudinal designs, would provide deeper insights into the long-term impact of retention strategies.

Doi : https://doi.org/10.21608/jaiep.2025.431657.1030
May 22, 2025; Revised: June 29, 2025; Accepted: July 8, 2025

Moreover, future research should explore the integration of technology, artificial intelligence (AI), and automation in predicting and mitigating employee turnover. Examining the effects of external factors, such as economic conditions and labor market dynamics, would also contribute to a more comprehensive understanding of the forces shaping employee retention.

References

1. Alam, A., & Asim, M. (2019). Relationship between job satisfaction and turnover intention. *International Journal of Human Resource Studies*, *9*(2), 163.
2. Alexander, P. M. (2020). *Strategies to Reduce Voluntary Employee Turnover* (Doctoral dissertation, Walden University).
3. Almerri, H.S.H. (2023). Investigating the impact of organizational culture on employee retention: Moderating role of employee engagement. Journal of System and Management Sciences, 13(4), 488–507.
4. An, S. H. (2019). Employee voluntary and involuntary turnover and organizational performance: Revisiting the hypothesis from classical public administration. *International Public Management Journal*, *22*(3), 444-469.
5. Chatwick, J. A. (2023). The Application of Information Technology to Human Resources Management in a Large Organizational Setting: An Examination of Concerns Related to Recruiting, Training, Retention, and Operational Efficiency (Doctoral dissertation, Trident University International).
6. El-Rayes, N., Fang, M., Smith, M., & Taylor, S. (2020). Predicting employee attrition using tree-based models. International Journal of Organizational Analysis, 28(6), 1273–1291.
7. Kurdi, B., & Alshurideh, M. (2020). Employee retention and organizational performance: Evidence from the banking industry. Management Science Letters, 10(16), 3981–3990.
8. Lee, Y., Yang, J., & Lee, J.Y. (2023). Identifying determinants in the first turnover of new graduate employees in Korea. Evidence-based HRM, 11(4), 641–658.
9. Madariaga, R., Oller, R., & Martori, J. C. (2018). Discrete choice and survival models in employee turnover analysis. Employee Relations, 40(2), 381-395.
10. Nasejje, J. B., Mwambi, H., Dheda, K., & Lesosky, M. (2017). A comparison of the conditional inference survival forest model to random survival forests based on a simulation study as well as on two applications with time-to-event data. *BMC medical research methodology*, *17*, 1-17.
11. Rahimi, B., Seproo, F. G., Khalilnezhad, R., Khodadost, M., Kashkalani, T., Younesi, F., ... & Bayat, M. (2024). Factors associated with the non-retention of health sector employees recruited in Iran: a survival analysis. BMC Health Services Research, 24(1), 942.