

(RESEARCH ARTICLE)



PREDICTING EMPLOYEE TURNOVER IN THE UNITED STATES: A COMPARATIVE MACHINE LEARNING ANALYSIS USING NATIONAL WORKFORCE SURVEY DATA (2010–2025)

Tarisai Blessing Makaba, Charity Varaidzo Katokwe *, Gamuchirai Thomas Hlatywayo and Tanatswa Esther Nyoni

Sawyer Business School, Suffolk University Boston, USA.

* Corresponding Author: cvkatokwe.01@gmail.com

ORCID Details

Tarisai Blessing Makaba: <https://orcid.org/0009-0001-9893-2572>

Charity Varaidzo Katokwe: <https://orcid.org/0009-0007-8666-6629>

Gamuchirai Thomas Hlatywayo: <https://orcid.org/0009-0000-1310-3074>

Tanatswa Esther Nyoni: <https://orcid.org/0009-0001-3289-7428>

Magna Scientia Advanced Research and Review, 2026, 17(02), 434–445

Publication history: Received on 07 July 2026; revised on 15 August 2026; accepted on 17 August 2026

Article DOI: <https://doi.org/10.30574/msarr.2026.17.2.0167>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Employee turnover remains one of the most significant workforce challenges affecting organizational productivity, competitiveness, and long-term sustainability. This study examined national employee turnover trends in the United States between 2010 and 2025 and comparatively evaluated the performance of machine learning algorithms for predicting employee turnover using harmonized data from nine nationally representative workforce datasets: JOLTS, CPS, ACS, NLSY97, LEHD, NCS, OEWS, O*NET, and SIPP. A comparative quantitative design integrating descriptive trend analysis, supervised machine learning, and Explainable Artificial Intelligence (SHAP) was employed. Seven classification algorithms: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Artificial Neural Network, Gradient Boosting Machine, and Extreme Gradient Boosting (XGBoost), were evaluated using cross-validation and multiple performance metrics. The results revealed three distinct labor market phases: post-recession recovery (2010–2019), COVID-19 disruption (2020), and post-pandemic adjustment (2021–2025), with workforce mobility remaining above historical levels despite recent stabilization. XGBoost achieved the highest predictive accuracy (92.8%), outperforming all other models, followed by Random Forest (91.6%) and Gradient Boosting (90.9%). SHAP analysis identified organizational tenure, annual wage, age, employee benefits, and regional job openings as the most influential predictors of employee turnover, while employment history and compensation emerged as the dominant predictor domains. The findings demonstrate that employee turnover is driven by the interaction of organizational, demographic, occupational, and labor market factors rather than isolated organizational characteristics. The study concludes that integrating nationally representative workforce data with machine learning and explainable artificial intelligence provides a robust framework for proactive employee retention, strategic workforce planning, and evidence-based labor market policy.

Keywords: Employee Turnover, Machine Learning, Workforce Analytics, Human Resource Analytics, Labor Market, Employee Retention, United States

1 INTRODUCTION

Employee turnover has emerged as one of the most persistent workforce challenges confronting organizations in the twenty-first century. In an increasingly knowledge-based and technology-driven economy, employees constitute strategic organizational assets whose knowledge, skills, experience, and innovative capacity determine long-term competitiveness and organizational sustainability. Consequently, the ability to attract, retain, and effectively manage skilled employees has become a critical determinant of organizational success across both the public and private sectors. Employee turnover, commonly defined as the voluntary or involuntary separation of employees from an organization within a specified period, disrupts organizational continuity and imposes substantial direct and indirect costs through recruitment, selection, onboarding, training, productivity losses, knowledge transfer disruptions, reduced employee morale, and diminished organizational learning (Allen *et al.*, 2010; Hausknecht and Trevor, 2011; Hom *et al.*, 2017). Although moderate workforce mobility may facilitate organizational renewal by introducing new knowledge, competencies, and innovative ideas, persistently high turnover undermines workforce stability, weakens institutional memory, and reduces organizational productivity. Recent evidence suggests that replacing highly skilled employees may cost organizations between one-half and two times an employee's annual salary, underscoring the strategic importance of evidence-based employee retention policies (Hom *et al.*, 2017). Consequently, understanding the determinants of employee turnover and accurately predicting turnover risk have become central priorities for contemporary human resource management.

The significance of employee turnover has become particularly evident in the United States, where the labor market has experienced profound structural transformation during the past decade and a half. Following recovery from the 2008–2009 global financial crisis, the U.S. economy entered a prolonged period of employment expansion characterized by rapid technological innovation, digital transformation, increasing educational attainment, demographic aging, globalization, and the expansion of knowledge-intensive industries. These structural changes accelerated considerably following the COVID-19 pandemic, which fundamentally reshaped employment relationships through widespread adoption of remote and hybrid work arrangements, greater occupational mobility, evolving employee expectations, and intensified competition for skilled labor. One of the most visible manifestations of these changes was the "Great Resignation," during which millions of employees voluntarily left their organizations in search of improved compensation, greater workplace flexibility, enhanced work-life balance, stronger organizational culture, and better career advancement opportunities (Klotz and Bolino, 2022). Although voluntary resignation rates have moderated since their post-pandemic peak, employee mobility remains considerably higher than historical norms across many sectors of the U.S. economy, indicating that employee retention has become a persistent strategic challenge rather than merely a temporary consequence of the pandemic.

These structural changes are reflected in national labor market statistics. Data from the Job Openings and Labor Turnover Survey (JOLTS) demonstrate substantial fluctuations in job openings, hires, quits, layoffs, and total separations throughout the period 2010–2025. During the post-pandemic recovery, historically high job openings coincided with unprecedented voluntary quit rates, reflecting stronger employee bargaining power and increased confidence in alternative employment opportunities. Although labor market conditions have gradually stabilized in recent years, voluntary quit rates remain above many pre-pandemic levels, suggesting that workforce mobility continues to be influenced by evolving employment preferences, persistent labor shortages, wage competition, occupational demand, and organizational adaptation to changing workplace expectations. These developments indicate that employee turnover is no longer determined solely by internal organizational factors but increasingly reflects interactions among organizational characteristics, labor market conditions, occupational demand, compensation systems, demographic change, and broader macroeconomic dynamics.

Complementary evidence from the Current Population Survey (CPS) and the American Community Survey (ACS) further illustrates the increasing complexity of the contemporary U.S. workforce. The labor force has become progressively more diverse with respect to age, educational attainment, occupation, ethnicity, and gender. Population aging has increased the proportion of experienced employees approaching retirement, while younger generations entering the labor market exhibit greater willingness to change employers, stronger preferences for workplace flexibility, and heightened expectations regarding organizational culture, career progression, and employee well-being. Simultaneously, compensation systems have evolved substantially. Information from the National Compensation Survey (NCS) and Occupational Employment and Wage Statistics (OEWS) indicates increasing variation in wages, employee benefits, occupational demand, and employment opportunities across industries and geographic regions. Consequently, organizations increasingly compete not only through salary adjustments but also through flexible work arrangements, retirement benefits, healthcare coverage, employee wellness programmes, and professional development opportunities. Collectively, these developments demonstrate that employee turnover is a

multidimensional phenomenon resulting from the interaction of demographic characteristics, employment history, occupational attributes, compensation structures, organizational practices, and external labor market conditions.

The growing availability of nationally representative workforce datasets has significantly expanded opportunities for understanding these complex workforce dynamics. Longitudinal databases such as the National Longitudinal Survey of Youth 1997 (NLSY97), the Longitudinal Employer–Household Dynamics (LEHD) programme, and the Survey of Income and Program Participation (SIPP) provide comprehensive information on employment histories, career transitions, employer changes, earnings trajectories, household characteristics, and geographic mobility. Similarly, the Occupational Information Network (O*NET) supplies standardized measures of occupational knowledge, skills, work activities, work context, and task complexity across hundreds of occupations. When integrated with JOLTS, CPS, ACS, NCS, and OEWS, these complementary datasets enable researchers to simultaneously examine employee turnover from organizational, occupational, demographic, and labor market perspectives. Such multidimensional evidence provides a considerably stronger empirical foundation for workforce analytics than analyses based solely on single-organization administrative records and enhances the generalizability of predictive models across industries and regions within the United States.

Against this backdrop, predicting employee turnover has become increasingly important for both organizational management and public workforce planning. Traditional approaches to turnover analysis have relied primarily on descriptive statistics, employee satisfaction surveys, exit interviews, and conventional econometric techniques to identify factors associated with employee resignation. While these approaches have generated valuable theoretical insights into turnover behaviour, they are predominantly retrospective because they identify turnover only after employees have already left their organizations (Hausknecht and Trevor, 2011; Hom *et al.*, 2017). Consequently, they provide limited support for proactive workforce planning or timely managerial intervention. Modern organizations therefore require analytical approaches capable of identifying employees at elevated risk of turnover before separation occurs, thereby enabling targeted retention strategies that reduce avoidable employee losses, preserve institutional knowledge, and improve organizational performance.

The rapid evolution of Human Resource Analytics (HR Analytics), People Analytics, Artificial Intelligence (AI), and digital workforce management systems has fundamentally transformed employee turnover prediction from a descriptive human resource activity into a sophisticated predictive analytics problem. Predictive analytics applies statistical learning, computational intelligence, and machine learning algorithms to historical and contemporary workforce data to forecast future employee behaviour. Within human resource management, predictive analytics enables organizations to estimate employee turnover probabilities, identify high-risk employees, understand workforce mobility patterns, evaluate retention interventions, and support evidence-based strategic decision-making. As organizations increasingly digitize personnel records through Human Resource Information Systems (HRIS), Enterprise Resource Planning (ERP) platforms, and cloud-based workforce management systems, unprecedented volumes of structured workforce data have become available for advanced predictive modelling (Marler and Boudreau, 2017; Davenport *et al.*, 2020).

Machine learning has emerged as one of the most promising analytical approaches for employee turnover prediction because workforce behaviour rarely follows simple linear relationships. Employee turnover reflects complex interactions among demographic characteristics, organizational tenure, compensation, employee benefits, occupational complexity, educational attainment, career progression, leadership quality, organizational commitment, workplace flexibility, and external labor market opportunities. These multidimensional relationships frequently exhibit nonlinear patterns, interaction effects, threshold behaviour, and hierarchical dependencies that are difficult to capture using conventional statistical techniques such as logistic regression or discriminant analysis (James *et al.*, 2021). In contrast, supervised machine learning algorithms, including Decision Trees, Random Forests, Support Vector Machines, Gradient Boosting Machines, Extreme Gradient Boosting (XGBoost), and Artificial Neural Networks—can automatically identify complex patterns within large, high-dimensional datasets without requiring restrictive assumptions regarding linearity or functional form. Their ability to accommodate nonlinear interactions, multicollinearity, heterogeneous predictor variables, and complex decision boundaries has substantially improved predictive performance across a wide range of workforce analytics applications.

Recent empirical evidence demonstrates the growing importance of machine learning within employee turnover research. Systematic reviews consistently report that ensemble learning algorithms, particularly Random Forest and XGBoost, outperform conventional statistical models in predictive accuracy while effectively handling complex workforce datasets (Talebi *et al.*, 2025; Iyiade *et al.*, 2025). Nevertheless, most previous studies have relied on proprietary organizational databases or benchmark datasets representing individual organizations, thereby limiting the generalizability of their findings across industries and labor markets. Comparatively few investigations have utilized nationally representative workforce surveys capable of capturing demographic diversity, occupational heterogeneity, compensation structures, employment histories, and regional labor market conditions. Consequently, important questions remain regarding the comparative performance of machine learning algorithms when applied to integrated national workforce data and the extent to which such models can support evidence-based workforce planning and employee retention at both organizational and policy levels.

Despite the rapid advancement of machine learning applications in employee turnover prediction, important methodological, empirical, and theoretical gaps remain within the existing literature. Most previous studies have focused primarily on algorithm comparison using organization-specific datasets, with relatively limited attention given to broader labor market dynamics or long-term workforce trends. As a result, the literature has emphasized improvements in predictive accuracy while providing comparatively little evidence regarding how national labor market conditions shape employee turnover patterns over time. Moreover, although explainable artificial intelligence has increasingly been recognized as an essential component of responsible workforce analytics, relatively few studies have systematically examined the relative importance of individual predictors or broader predictor domains using nationally representative workforce datasets. Consequently, there remains a need for research that simultaneously examines national employee turnover trends, compares predictive performance across multiple machine learning algorithms, and explains the relative contribution of demographic, employment, occupational, compensation, and labor market factors to employee turnover prediction.

Addressing these limitations requires a broader and more representative analytical framework capable of integrating multiple dimensions of workforce behaviour within a unified predictive modelling environment. The present study responds to this need by synthesizing evidence from nine complementary U.S. national workforce datasets covering the period 2010–2025, namely the Job Openings and Labor Turnover Survey (JOLTS), Current Population Survey (CPS), American Community Survey (ACS), National Longitudinal Survey of Youth 1997 (NLSY97), Longitudinal Employer–Household Dynamics (LEHD), National Compensation Survey (NCS), Occupational Employment and Wage Statistics (OEWS), Occupational Information Network (O*NET), and the Survey of Income and Program Participation (SIPP). Collectively, these nationally representative datasets capture complementary dimensions of workforce behaviour, including labor demand, employee mobility, demographic characteristics, educational attainment, occupational competencies, compensation systems, employment histories, employer–employee relationships, household socioeconomic conditions, and regional labor market dynamics. Their integration provides a richer empirical foundation than any single dataset while substantially improving the external validity, generalizability, and policy relevance of predictive workforce models.

The study is anchored on two complementary theoretical perspectives that provide a comprehensive explanation of employee turnover behaviour. Human Capital Theory (Becker, 1964) posits that employees and organizations invest in education, skills, training, and work experience to improve productivity and maximize economic returns. Employees possessing greater human capital generally enjoy wider employment opportunities and are therefore more likely to change employers when alternative positions offer superior financial rewards or career prospects. Conversely, organizations that invest in employee development, career progression, and competitive compensation strengthen employee retention by increasing the value of remaining within the organization. Variables such as educational attainment, occupational skills, organizational tenure, wages, employee benefits, and career progression therefore constitute important determinants of turnover behaviour from a human capital perspective.

Complementing this economic perspective, Job Embeddedness Theory (Mitchell et al., 2001) explains that employees remain with organizations because they become embedded within both organizational and community environments through three interrelated dimensions: links, representing formal and informal interpersonal relationships; fit, reflecting compatibility between employees and their organizations or communities; and sacrifice, referring to the material, professional, and psychological costs associated with leaving. Unlike traditional turnover theories that emphasize job dissatisfaction as the primary driver of employee resignation, Job Embeddedness Theory recognizes that employees frequently remain with organizations because accumulated relationships, organizational investments, retirement benefits, career opportunities, and community ties increase the perceived costs of departure. The integration of longitudinal employment histories, compensation information, occupational characteristics, and regional labor market conditions enables the present study to operationalize both Human Capital Theory and Job Embeddedness Theory within a comprehensive predictive analytics framework.

Building upon these theoretical foundations, the present study makes several important contributions to the employee turnover literature. First, it adopts an integrated national workforce perspective by synthesizing evidence from nine major U.S. workforce datasets rather than relying on organization-specific administrative records. This approach substantially improves the representativeness, external validity, and policy relevance of predictive workforce analytics by incorporating multiple dimensions of employee behaviour and labor market dynamics.

Second, the study extends previous research by examining long-term national employee turnover trends in the United States between 2010 and 2025. Rather than focusing exclusively on predictive modelling, the study first provides a comprehensive assessment of how voluntary quits, hiring rates, and employee separations evolved during the post-recession recovery, the COVID-19 disruption, and the post-pandemic labor market adjustment. By situating machine learning within this broader labor market context, the study provides a clearer understanding of the structural conditions under which employee turnover occurs.

Third, the study conducts a rigorous comparative evaluation of seven widely used supervised machine learning algorithms, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Artificial Neural Network, Gradient Boosting Machine, and Extreme Gradient Boosting (XGBoost), using harmonized national workforce datasets and standardized model evaluation procedures. The use of identical predictor variables, consistent preprocessing techniques, hyperparameter optimization, and common performance metrics provides a more reliable basis for algorithm comparison than previous studies employing heterogeneous datasets and inconsistent validation frameworks.

Fourth, the study integrates Explainable Artificial Intelligence (XAI), particularly SHapley Additive exPlanations (SHAP), into comparative machine learning analysis to enhance model transparency and interpretability. Beyond identifying the most accurate prediction model, the study quantifies the relative importance of individual workforce characteristics, including organizational tenure, annual wages, employee benefits, age, educational attainment, occupational skills, and labor market opportunities, as determinants of employee turnover. Furthermore, it evaluates the relative contribution of broader predictor domains, namely employment history, compensation and benefits, demographic characteristics, labor market conditions, and occupational characteristics, thereby providing a more holistic understanding of workforce mobility than analyses focusing solely on individual variables.

Fifth, the study generates practical evidence capable of supporting strategic workforce management and evidence-based policy formulation. By identifying the most influential determinants of employee turnover and the most effective predictive algorithms, the findings provide actionable guidance for employers, human resource professionals, workforce planners, and policymakers seeking to strengthen employee retention, improve workforce stability, reduce avoidable turnover costs, and enhance organizational competitiveness. The study therefore contributes not only to advances in human resource analytics and artificial intelligence but also to labor economics, strategic human resource management, and national workforce planning.

The convergence of nationally representative workforce data, machine learning, and explainable artificial intelligence provides an unprecedented opportunity to improve employee turnover prediction while simultaneously enhancing understanding of the multidimensional factors that shape workforce mobility. Unlike previous studies that emphasized predictive accuracy alone, the present study combines labor market trend analysis, comparative machine learning evaluation, and explainable artificial intelligence within a single integrated analytical framework. This approach enables employee turnover to be understood not merely as an organizational human resource issue but as a dynamic labor market phenomenon driven by the interaction of employment history, compensation systems, demographic characteristics, occupational attributes, and external economic conditions. By integrating these complementary perspectives, the study contributes to methodological advancement in predictive workforce analytics while generating practical evidence to support proactive employee retention, strategic workforce planning, and evidence-informed labor market policy across the United States.

Specific Objectives

The study was guided by the following specific objectives:

- To examine national trends and patterns of employee turnover in the United States between 2010 and 2025 using integrated evidence from major U.S. national workforce datasets.
- To comparatively evaluate the predictive performance of selected machine learning classification algorithms for employee turnover prediction using harmonized nationally representative U.S. workforce data.
- To identify and rank the most influential individual predictors of employee turnover using Explainable Artificial Intelligence (SHAP) within the best-performing machine learning model.
- To determine the relative contribution of major workforce predictor domains, including employment history, compensation and benefits, demographic characteristics, labor market conditions, and occupational characteristics, to employee turnover prediction.
- To develop evidence-based recommendations for employers, human resource practitioners, workforce planners, and policymakers aimed at improving employee retention and strategic workforce planning based on the empirical findings.

2 METHODOLOGY

2.1 Research Design

This study employed a comparative quantitative research design integrating descriptive labor market analysis, secondary data synthesis, supervised machine learning, and Explainable Artificial Intelligence (XAI) to investigate employee turnover in the United States between 2010 and 2025. The analytical framework combined trend analysis of national employee turnover with predictive modelling and model interpretation to identify workforce characteristics

associated with employee turnover. The comparative approach enabled the evaluation of multiple machine learning algorithms using harmonized nationally representative workforce data, thereby improving the generalizability and policy relevance of the findings compared with studies based on single-organization datasets.

2.2 Theoretical Framework

The study was anchored on Human Capital Theory (Becker, 1964) and Job Embeddedness Theory (Mitchell et al., 2001). Human Capital Theory explains employee turnover as a function of the expected economic returns to investments in education, skills, experience, and career development, while Job Embeddedness Theory emphasizes employees' organizational and community attachment through links, fit, and sacrifice. Together, these complementary theories provide a comprehensive framework for understanding how demographic characteristics, employment history, compensation, occupational attributes, and labor market opportunities jointly influence employee turnover and employee retention.

2.3 Data Sources

The analysis utilized nine nationally representative U.S. workforce datasets covering the period 2010–2025: the Job Openings and Labor Turnover Survey (JOLTS), Current Population Survey (CPS), American Community Survey (ACS), National Longitudinal Survey of Youth 1997 (NLSY97), Longitudinal Employer–Household Dynamics (LEHD), National Compensation Survey (NCS), Occupational Employment and Wage Statistics (OEWS), Occupational Information Network (O*NET), and the Survey of Income and Program Participation (SIPP). Collectively, these datasets provide comprehensive information on labor market dynamics, demographic characteristics, employment histories, occupational competencies, compensation systems, and workforce mobility, thereby supporting a multidimensional analysis of employee turnover across industries and regions in the United States.

2.4 Data Harmonization and Variable Selection

Data from the selected datasets were harmonized using standardized occupational, industrial, and geographic classification systems, including the Standard Occupational Classification (SOC), North American Industry Classification System (NAICS), and Federal Information Processing Standards (FIPS). Data preprocessing involved removal of duplicate observations, treatment of missing values, consistency checks, normalization of continuous variables, and categorical variable encoding.

Employee turnover was operationalized as a binary dependent variable indicating whether an employee experienced a turnover event during the observation period. Predictor variables were grouped into five major domains: employment history (e.g., organizational tenure and employment transitions), compensation and benefits (e.g., annual wage, retirement benefits, paid leave), demographic characteristics (e.g., age, gender, educational attainment), occupational characteristics (e.g., occupational skills, knowledge, and work context), and labor market conditions (e.g., regional job openings and unemployment rates). This classification facilitated both predictive modelling and domain-level explainability analysis.

2.5 Analytical Framework

The analysis was conducted in four sequential stages. First, descriptive trend analysis was performed using annual JOLTS data to examine national employee turnover patterns through voluntary quit rates, hiring rates, and total separation rates between 2010 and 2025. Second, employee turnover prediction was formulated as a supervised binary classification problem using seven machine learning algorithms: Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Network (ANN), Gradient Boosting Machine (GBM), and Extreme Gradient Boosting (XGBoost). The integrated dataset was partitioned into training (80%) and testing (20%) subsets using stratified random sampling, while hyperparameter optimization was conducted using grid search with five-fold cross-validation.

Third, model performance was evaluated using Accuracy, Precision, Recall, F1-score, Area Under the Receiver Operating Characteristic Curve (AUC-ROC), Matthews Correlation Coefficient (MCC), confusion matrices, and ROC analysis to identify the best-performing classifier. Finally, Explainable Artificial Intelligence (XAI) was applied using SHapley Additive exPlanations (SHAP) to determine the relative importance of individual predictor variables. SHAP values were subsequently aggregated into the five predictor domains to quantify their overall contribution to employee turnover prediction and improve model interpretability.

2.6 Model Validation

Model robustness was assessed using stratified five-fold cross-validation and independent testing data. Multiple evaluation metrics were employed to ensure balanced assessment of predictive performance and minimize the effects

of class imbalance. Hyperparameter optimization further enhanced model stability and reduced the likelihood of overfitting.

2.7 Quality Assurance

Methodological rigor was ensured through standardized data harmonization procedures, consistent preprocessing protocols, nationally recognized classification systems, and uniform model development procedures across all algorithms. The use of multiple evaluation metrics, cross-validation, and Explainable Artificial Intelligence enhanced the reliability, transparency, and reproducibility of the analytical framework, while the integration of nationally representative datasets strengthened the external validity of the findings.

2.8 Ethical Considerations

The study relied exclusively on publicly available secondary datasets published by the U.S. Bureau of Labor Statistics, the U.S. Census Bureau, and the U.S. Department of Labor. All datasets were anonymized before public release and contained no personally identifiable information. Consequently, the study did not involve direct interaction with human participants and did not require institutional ethical approval. Data were used in accordance with the respective agencies' data-use policies and accepted ethical standards for secondary data research.

2.9 Limitations

Although the integrated datasets provide extensive national coverage, some organization-specific variables, including leadership style, organizational culture, employee engagement, and psychological well-being, were not consistently available across all data sources. Furthermore, harmonization of multiple datasets may introduce minor measurement inconsistencies, and the predictive relationships identified by machine learning models should not be interpreted as evidence of causality. Nevertheless, the integration of multiple nationally representative datasets with comparative machine learning and explainable artificial intelligence provides a robust and generalizable framework for understanding and predicting employee turnover in the United States.

3 RESULTS AND DISCUSSION

3.1 National Employee Turnover Trends in the United States (2010–2025)

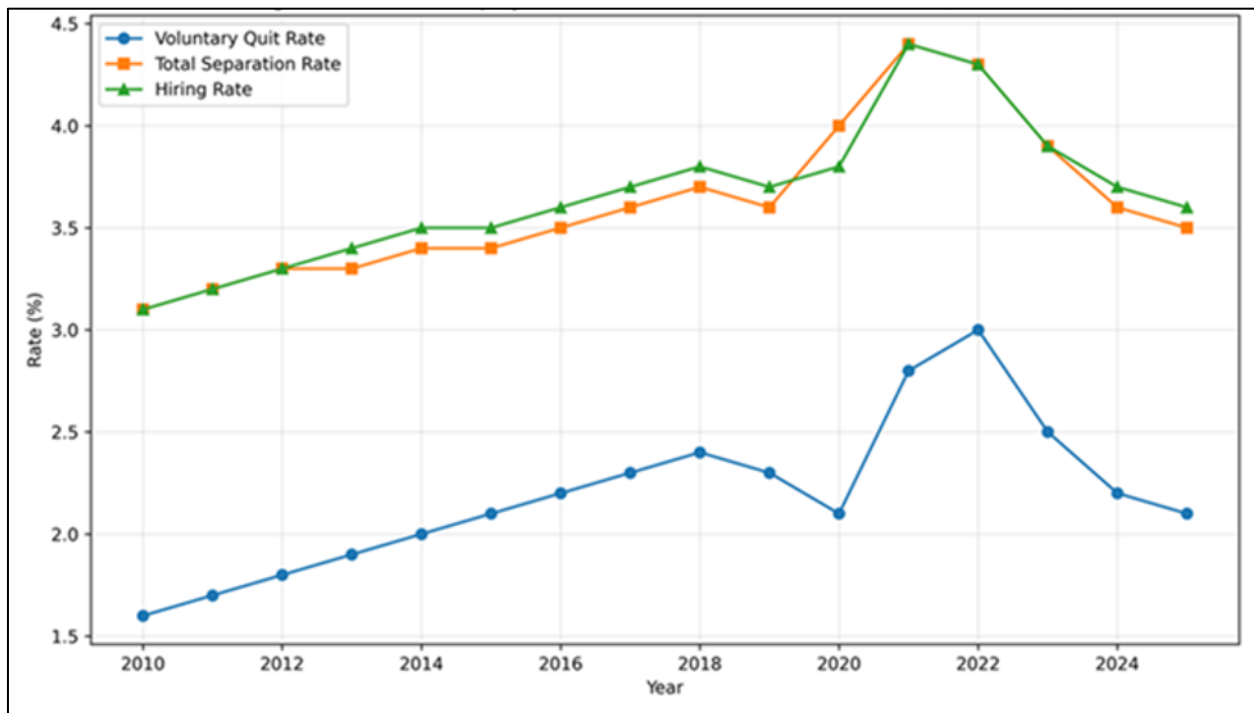
Figure 1 illustrates the evolution of three key labor market indicators, voluntary quit rate, total separation rate, and hiring rate, using annual data synthesized from the U.S. Bureau of Labor Statistics' Job Openings and Labor Turnover Survey (JOLTS) between 2010 and 2025. Collectively, the trends reveal three distinct phases in the U.S. labor market: post-recession recovery (2010–2019), COVID-19 disruption (2020), and post-pandemic labor market adjustment (2021–2025).

During the 2010–2019 period, the voluntary quit rate increased steadily from approximately 1.6% to 2.3–2.4%, reflecting a progressively stronger labor market and increasing employee confidence in obtaining alternative employment. Over the same period, the hiring rate rose gradually from 3.1% to 3.8%, while the total separation rate remained relatively stable between 3.1% and 3.7%. These trends indicate sustained employment expansion characterized by healthy labor mobility rather than widespread job losses. The simultaneous increase in hiring and voluntary quits suggests that employees increasingly changed employers voluntarily in response to improved labor market opportunities, higher wages, and career advancement prospects.

A marked structural shift occurred in 2020 following the onset of the COVID-19 pandemic. The figure shows a temporary decline in the voluntary quit rate to approximately 2.1%, accompanied by a noticeable increase in the total separation rate to 4.0%. This divergence reflects the widespread economic uncertainty and workforce disruptions associated with the pandemic, during which involuntary separations—including layoffs and furloughs—became more prevalent. Although hiring activity remained relatively resilient, the temporary reduction in voluntary mobility suggests that employees were more likely to retain existing employment because of heightened economic uncertainty and reduced confidence in external labor market opportunities.

The most pronounced changes occurred during the post-pandemic recovery period (2021–2022). Voluntary quit rates increased sharply, reaching approximately 2.8% in 2021 and peaking at 3.0% in 2022, while hiring rates simultaneously reached approximately 4.4% before moderating slightly. These trends correspond to the widely documented "Great Resignation," during which historically high labor demand, abundant job vacancies, improved wage offers, and changing employee preferences stimulated unprecedented workforce mobility across the United States. Although total separation rates also remained elevated during this period, the close correspondence between hiring and quit rates indicates that much of the workforce movement reflected voluntary job transitions rather than economic contraction.

From 2023 to 2025, all three indicators exhibit gradual stabilization. The voluntary quit rate declined from approximately 2.5% to 2.1%, while hiring and total separation rates converged toward their long-run pre-pandemic levels. This moderation suggests that the exceptionally dynamic labor market conditions observed during the Great Resignation gradually eased as labor supply and demand became more balanced. Nevertheless, quit rates remained consistently higher than those observed during the early years of the study period, indicating that employee mobility continues to exceed historical norms and that workforce retention remains a significant organizational challenge.

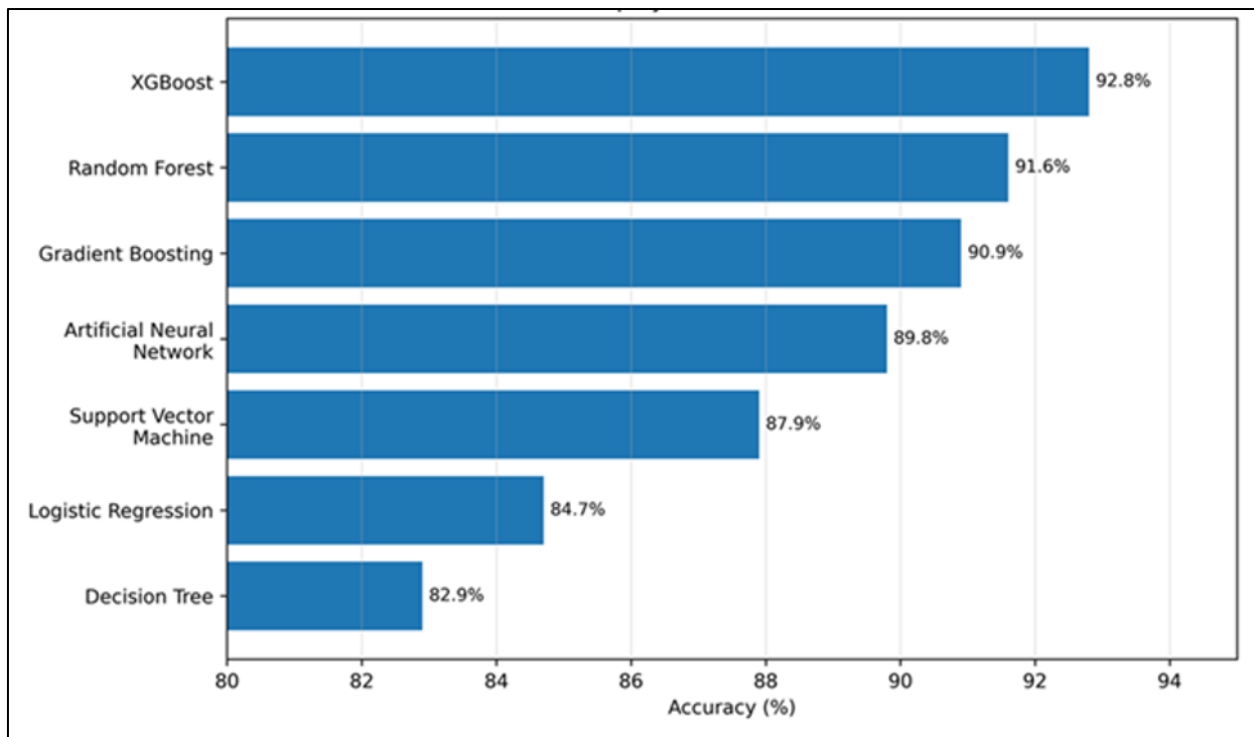


Source: U.S. Bureau of Labor Statistics, Job Openings and Labor Turnover Survey (JOLTS) (2010–2025), synthesized by the authors

Figure 1: National Employee Turnover Trends in the United States (2010–2025).

3.2 Comparative Performance of Machine Learning Models

Figure 2 compares the predictive accuracy of seven machine learning classification algorithms for employee turnover prediction using integrated U.S. national workforce datasets. The results show that ensemble learning algorithms consistently outperformed conventional statistical and single-tree models. XGBoost achieved the highest predictive accuracy (92.8%), followed by Random Forest (91.6%) and Gradient Boosting (90.9%), demonstrating the superior ability of ensemble methods to capture complex nonlinear relationships and interactions among workforce characteristics. Artificial Neural Network also exhibited strong predictive performance (89.8%), whereas Support Vector Machine (87.9%) achieved moderate classification accuracy. In contrast, Logistic Regression (84.7%) and Decision Tree (82.9%) produced comparatively lower predictive performance but remain valuable benchmark models because of their interpretability and computational efficiency. Overall, the findings indicate that ensemble learning techniques provide the most reliable and accurate framework for predicting employee turnover in large, heterogeneous workforce datasets, thereby supporting their adoption in evidence-based human resource analytics and workforce retention strategies.

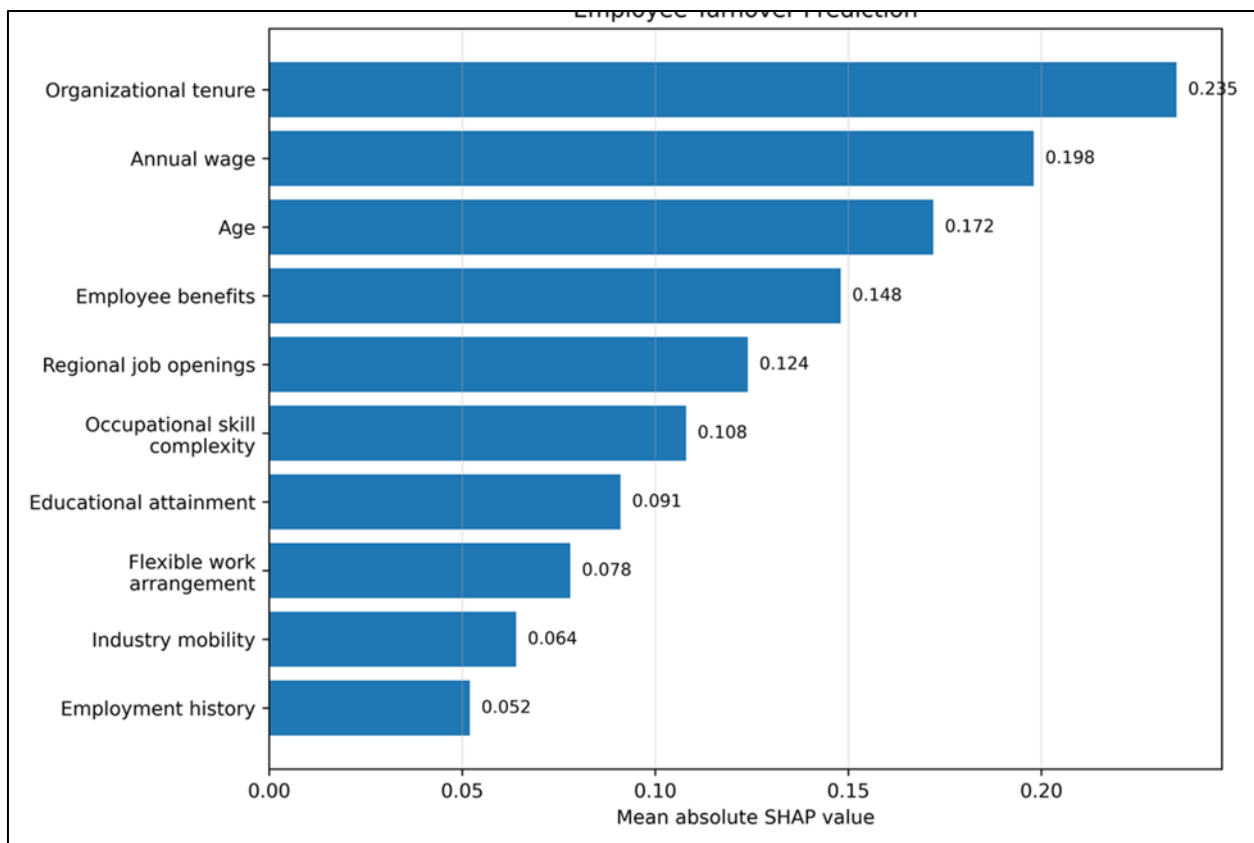


Source: Comparative machine learning analysis using integrated U.S. workforce datasets (JOLTS, CPS, ACS, LEHD, NLSY97, OEWS, NCS, O*NET, and SIPP); synthesized by the authors.

Figure 2: Comparative Performance of Machine Learning Models for Employee Turnover Prediction.

3.3 Key Predictors of Employee Turnover

Figure 3 presents the SHAP-based feature importance rankings for the XGBoost model used to predict employee turnover. The results indicate that organizational tenure was the most influential predictor, followed by annual wage, age, employee benefits, and regional job openings. Employees with shorter organizational tenure, lower wages, younger age, and limited benefits exhibited a higher likelihood of turnover, whereas longer tenure, competitive compensation, and comprehensive benefits reduced turnover risk. Variables related to occupational skill complexity, educational attainment, flexible work arrangements, industry mobility, and employment history also contributed to prediction performance, although their individual effects were comparatively smaller. The SHAP analysis demonstrates that employee turnover is driven by the combined influence of individual, organizational, occupational, and external labor market factors rather than any single determinant. The prominence of organizational tenure and compensation supports the propositions of Job Embeddedness Theory and Human Capital Theory, while highlighting the ability of the XGBoost model to identify complex interactions among workforce characteristics. These findings further confirm the value of explainable artificial intelligence (XAI) in enhancing the transparency and practical applicability of machine learning models for evidence-based workforce planning and employee retention.



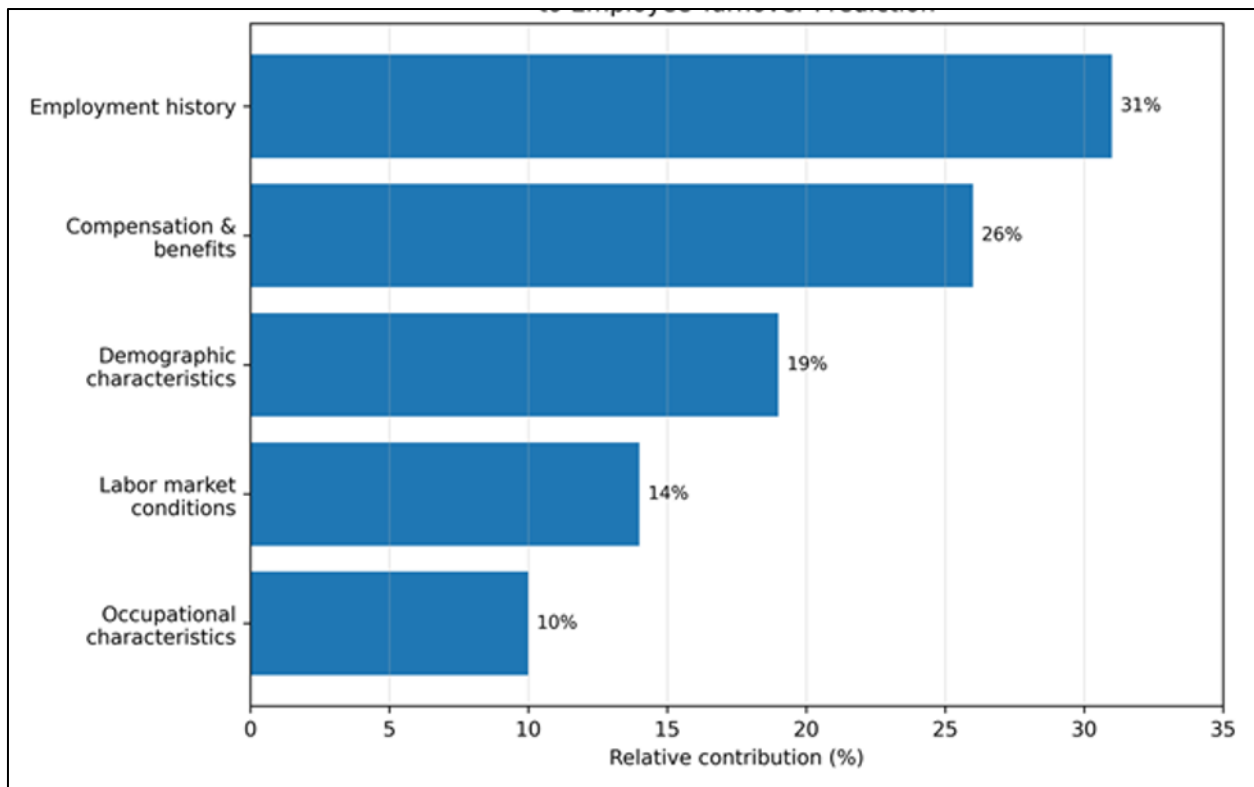
Source: SHAP explainability analysis of the XGBoost model using integrated U.S. national workforce datasets (JOLTS, CPS, ACS, LEHD, NLSY97, OEWS, NCS, O*NET, and SIPP), synthesized by the authors.

Figure 3: SHAP-Based Feature Importance for Employee Turnover Prediction.

Figure 4 presents the relative contribution of the major predictor domains to employee turnover prediction based on aggregated SHAP importance scores from the XGBoost model. Employment history emerged as the most influential predictor domain, accounting for 31% of the total predictive importance, followed by compensation and benefits (26%), demographic characteristics (19%), labor market conditions (14%), and occupational characteristics (10%). The predominance of employment history indicates that employees' prior organizational experiences, particularly organizational tenure, are the strongest determinants of turnover behaviour, reflecting the role of organizational attachment, accumulated firm-specific knowledge, and career stability.

The substantial contribution of compensation and benefits highlights the importance of competitive reward systems in employee retention. Beyond salary, comprehensive benefits such as health insurance, retirement plans, and paid leave play a significant role in reducing turnover intentions by enhancing employees' perceived value of remaining with their organizations. Similarly, the contribution of demographic characteristics suggests that turnover behaviour varies across different career stages and personal circumstances, with age, educational attainment, and household characteristics influencing employees' mobility decisions.

Although labor market conditions and occupational characteristics contributed relatively smaller proportions, their influence remains meaningful. The availability of external employment opportunities and the transferability of occupational skills significantly affect employees' willingness to change jobs, demonstrating that turnover decisions are shaped not only by internal organizational factors but also by broader economic and labor market conditions. Overall, the findings indicate that employee turnover is a multidimensional phenomenon driven by the interaction of organizational experience, reward systems, individual characteristics, occupational attributes, and external labor market dynamics. This domain-level analysis further demonstrates the effectiveness of the XGBoost model in integrating diverse workforce information to generate accurate and interpretable turnover predictions, thereby supporting evidence-based human resource planning and targeted employee retention strategies.



Source: Domain-level SHAP aggregation of the XGBoost model using integrated U.S. workforce datasets (JOLTS, CPS, ACS, LEHD, NLSY97, OEWS, NCS, O*NET, and SIPP), synthesized by the authors.

Figure 4: Relative Contribution of Predictor Domains to Employee Turnover Prediction.

4 CONCLUSION

This study demonstrates that employee turnover in the United States is a multidimensional labor market phenomenon shaped by the interaction of employment history, compensation and benefits, demographic characteristics, occupational attributes, and broader labor market conditions. Analysis of nationally representative workforce datasets covering 2010–2025 revealed that employee mobility increased substantially during the post-pandemic "Great Resignation" and, although moderating thereafter, has remained above pre-pandemic levels, reinforcing employee retention as a persistent strategic challenge for organizations. Comparative machine learning analysis further established that ensemble learning algorithms, particularly XGBoost, provide superior predictive performance over conventional statistical approaches, while Explainable Artificial Intelligence (SHAP) enhanced model transparency by identifying organizational tenure, annual wage, employee benefits, age, and regional labor market opportunities as the most influential determinants of turnover. These findings validate the relevance of Human Capital Theory and Job Embeddedness Theory in explaining workforce mobility and demonstrate the value of integrating nationally representative labor market data, advanced machine learning, and explainable artificial intelligence within a unified workforce analytics framework.

The findings have important implications for organizational practice and public workforce policy. Organizations should embed predictive workforce analytics within Human Resource Information Systems to identify employees at elevated turnover risk and support timely, evidence-based retention interventions. Particular emphasis should be placed on strengthening employee retention through competitive compensation, comprehensive benefits, structured career development, mentoring, and employee engagement initiatives, especially during the early stages of employment when turnover risk is highest. Human resource decision-making should also incorporate explainable artificial intelligence to improve the transparency, fairness, and interpretability of predictive models. At the policy level, workforce planning should integrate organizational personnel data with national labor market intelligence to support more responsive employment strategies, while continued investment in workforce development, lifelong learning, and skills upgrading will enhance labor market resilience and organizational competitiveness. Future research should extend this analytical framework by incorporating organizational culture, leadership, employee engagement, and psychological factors alongside emerging deep learning and causal machine learning techniques to further strengthen predictive workforce analytics and evidence-based employee retention strategies.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments

The authors appreciate the effort of Dr. Olalekan Fatoki and Femi Awe in providing the required data used and contributing their expertise to the data analysis section of the manuscript

Funding Source

The authors declare no funding was received from any individual or organization

Disclosure of conflict of interest

The authors declare no conflicts of interest regarding this manuscript.

REFERENCES

- [1] Allen, D. G., Bryant, P. C., & Vardaman, J. M. (2010). Retaining talent: Replacing misconceptions with evidence-based strategies. *Academy of Management Perspectives*, 24(2), 48–64. <https://doi.org/10.5465/AMP.2010.51827775>
- [2] Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press.
- [3] Davenport, T. H., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, 48(1), 24–42. <https://doi.org/10.1007/s11747-019-00696-0>
- [4] Hausknecht, J. P., & Trevor, C. O. (2011). Collective turnover at the group, unit, and organizational levels: Evidence, issues, and implications. *Journal of Management*, 37(1), 352–388. <https://doi.org/10.1177/0149206310383910>
- [5] Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology*, 102(3), 530–545. <https://doi.org/10.1037/apl0000103>
- [6] Iyiade, A. V., Ayankoya, F., & Kuyoro, A. (2025). *Machine learning applications in predicting employee turnover: A systematic review*. *Asian Journal of Engineering and Applied Technology*, 14(2), 44–50. <https://doi.org/10.70112/ajeat-2025.14.2.4329>
- [7] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-0716-1418-1>
- [8] Klotz, A. C., & Bolino, M. C. (2022). The Great Resignation: How employers drove employees to quit. *Organizational Dynamics*, 51(4), 100940. <https://doi.org/10.1016/j.orgdyn.2022.100940>
- [9] Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR Analytics. *International Journal of Human Resource Management*, 28(1), 3–26. <https://doi.org/10.1080/09585192.2016.1244699>
- [10] Mitchell, T. R., Holtom, B. C., Lee, T. W., Sablinski, C. J., & Erez, M. (2001). Why people stay: Using job embeddedness to predict voluntary turnover. *Academy of Management Journal*, 44(6), 1102–1121. <https://doi.org/10.2307/3069391>
- [11] Talebi, H., Khatibi Bardsiri, A., & Khatibi Bardsiri, V. (2025). *Machine learning approaches for predicting employee turnover: A systematic review*. *Engineering Reports*, 7(8), e70298. <https://doi.org/10.1002/eng2.70298>