



SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Explainable Artificial Intelligence for Intelligent Human Resource Analytics: Predicting Employee Retention and Workforce Performance

Prof. (Dr.) Sharmila Subramanian¹, Madhukumar PS², Zunzarrao Vilasrao Thorat³, Dr. S. Jyothirmaye Reddy⁴, Dr Chaitali Shewale⁵, Raj Kumar Singh⁶

¹Principal, ASTHA School of Management, Bhubaneswar, Plot no. 88/1 pvt plot CHP lane 3, Kanan Vihar phase one, P. O. Patia, Bhubaneswar. Odisha, PIN :- 751024, Email :- sharmila.su@gmail.com, Orcid Id: <https://orcid.org/0000-0001-6256-8687>

²Associate professor, Management, Xavier Institute of Management & Entrepreneurship, Bangalore, Email: madhukumar@xime.org, Orchid ID: 0000-0002-6017-119X

³Assistant Professor, Department: EXTC, Specialization: COMMUNICATION, University name with full address(city, pin code): Bharati Vidyapeeth College of Engineering Navi Mumbai, Email ID: zunzar_thorat@rediffmail.com

⁴Associate Professor, Faculty of Management Studies, HRM, University name with full address(city, pin code): Marwadi, University, Rajkot-Morbi Highway Rajkot- 360003, Gujarat, Email Id: sathi.reddy@marwadieducation.edu.in, Orcid Id:0000-0002-1629-5777

⁵Assistant professor, College: VIT, pune, IT department, Email id: chaitali.shewale@vit.edu, ORCID ID: 0000-0002-7293-1766

⁶Professor & HOD, Department: Deptt. of Commerce, College: School of Management Sciences, City / District/Pincode: Varanasi -221005, UP, India, ORCID ID: 0000-0003-1989-8714, Email ID: rksingh@smsvaranasi.com

ABSTRACT

While employee retention and workforce performance are pivotal to the continuity of organizations, there may be less clarity around the factors influencing the output of employees in predictive HR models. This study developed and explained machine-learning models for employee retention and workforce performance and assessed gender-based differences in retention-model performance. Two independent cross-sectional datasets comprising 1,191 employees for retention classification and 368 employees for performance prediction were analyzed. Logistic regression and random forest classification were evaluated for retention, while multiple linear regression, random forest regression, and a mean baseline were compared for performance. Models were validated using repeated five-fold cross-validation and out-of-fold predictions. Permutation importance and SHAP were applied for model explanation, followed by sensitivity and gender-subgroup assessments. The retention random forest achieved an ROC-AUC of 0.824 and balanced accuracy of 0.735. Years of experience was the strongest retention predictor, followed by job stability, deserved promotion, emotional commitment, and training participation. The performance of the random forest achieved an R-squared of 0.686 and a root mean squared error of 0.393. Employee readiness was the dominant performance predictor, followed by organizational support culture, innovative HR strategies, and digital HR technologies. Gender-specific ROC-AUC and balanced accuracy were similar, although predicted retention rates and class-specific errors differed. Explainable machine learning can complement workforce analytics with beneficial predictability and interpretable individual employee factors to enable transparent workforce analytics. There's a need for human involvement, external validation, and ongoing fairness monitoring in the responsible use of HR.

Keywords: employee retention, workforce performance, explainable artificial intelligence, human resource analytics, machine learning

1. Introduction

Employee retention and workforce performance go hand in hand with organizational continuity, productivity and competitive advantage. Organizations need to be able to keep the employees they have and meet the needs of members who can work effectively in a technologically challenging environment. When staff turnover occurs, it can cause instability in operations, cost organizations more to recruit and train employees, and reduce the knowledge gained from experience. HRM practices such as compensation, career growth, training, recognition, participation and supervisory support can enhance the employees' desire to stay with the organization (Azeez, 2017). It is now no longer an administrative issue but rather a strategic focus that has been influenced by the evolving expectations of employees, workforce mobility and global competition for skilled employees (Singh, 2019). These are personal experiences and conditions, as well as organizational conditions. Factors such as job satisfaction, commitment, perceived alternatives, professional growth, workplace relationships, and organizational support may impact turnover (Hom et al., 2019). Determining what the most effective predictors are of retention can help an organization become aware of which elements in the workforce are at risk for attrition and focus efforts on elements that are influential and feasible to intervene upon.

Another critical aspect of an organisation's efficiency is its workforce performance. It covers task achievement, ability to adapt, contextual contribution and the employees' ability to respond to changing demands of work. In the era of digital transformation, an organization's employees are expected to perform their duties as envisaged, together with adapting to new systems and revamped processes (Pradhan & Jena, 2017).

Additionally, workforce engagement is seen as a means of supporting performance by linking the cognitive, emotional and behavioural engagement of the workforce to organizational goals, service quality and operational performance (Schneider et al., 2018). The organizational climate in which the abilities are used, as well as the abilities themselves, thus have an impact on performance. This can be further enhanced through innovative human resource strategies that promote learning, participation, autonomy, collaboration and knowledge sharing. Research on HRM suggests that HRM practices that are designed properly can increase employees' innovative work behaviour through providing opportunities and motivation for them to have new ideas and implement them (Bos-Nehles et al., 2017). Also, the perceptions of HR practices among employees are important, as formal policies can have different effects on behaviour depending on the climate existing around the policy, in terms of experimentation and innovation (Bos-Nehles & Veenendaal, 2019). The results of this study place innovation in HR in a new light as a possible factor in workforce performance and highlight the value of organisational conditions.

Digital HR technologies have given organizations access to a wider range of workforce information, automated mundane tasks, and provided an opportunity to create a personalized workplace experience through greater employee information and support evidence-based decisions. Adoption also impacts the HR structure, professional roles, and relationships among employees and management (DiRomualdo et al., 2018). Digital transformation is not just about installing technology; it's about how the technology is implemented and whether the outcomes are embraced by employees, the organization, and integrated into HR processes (Zhang & Chen, 2024). The culture of the organization and whether its employees are ready for change may then be the deciding factor in whether innovative HR strategies and digital technologies drive improved performance. Machine learning can offer an appropriate analytic tool to analyse complex relationships in workforce data. Instead of the predefined linear relationships used in traditional models, machine-learning algorithms have the potential to reflect nonlinear relationships and interactions between many of these employee attributes. AI's use in workforce management has been shown to assess job performance, identify trends in employee behavior, and enhance HR processes (Ramachandran et al., 2022). However, in an employment context, predictive accuracy is not enough, as the opacity of the outputs could make it hard for management to trust and hard for them to assess the consequences of the recommendations. Explainable AI (XAI) overcomes this concern by providing insights into the contribution of the predictors at both a global and individual level. Explainability can help render complex models more transparent and make it possible to look at whether the predictions are based on meaningful and defensible information (Gunning et al., 2019). In the HR field, its application can assist in interpreting the predictions that are made related to employees and comprehend the basis of computations that are used behind the model's recommendations (Abonamah et al., 2022). It is also important to evaluate model assessment with regard to whether the overall predictive accuracy and/or error patterns vary between employee groups because this can mask group differences in overall accuracy.

While there is abundant research on these topics of employee retention, workforce performance, HR innovation, and digital transformation, these areas are typically studied independently. There is limited research that attempts to link the classification of retention with predictions of performance and then explains the model, tests model sensitivity and considers the model's fairness to subgroups. There is a limited amount of evidence in the Saudi context regarding explainable retention prediction and the relative impact of innovative HR practices, digital technologies, employee readiness, and organizational support on employee performance.

The objective of this study was to create and validate machine learning models to predict employee retention based on demographic, occupational, organizational, and attitudinal attributes; predict employee performance using innovative HR strategies, digital HR technologies, employee readiness, and organizational support culture; explain predictions made; and identify gender-based differences in the performance of retention models.

2. Methodology

2.1 Research Design

This study was designed in a quantitative, retrospective predictive-analytics approach to analyze employee retention and workforce performance. Two independent, cross-sectional employee data sets were analysed due to the lack of a common respondent-level identifier for the data sets. All of the models assumed that employee retention was a binary classification outcome and that workforce performance was a continuous outcome. Then, XAI approaches were used to determine the most influential variables for each prediction.

2.2 Data Sources

The employee-retention data were taken from an open survey of the employees in Saudi Arabia (Alqahtani et al., 2025). There were 1,191 respondents in the dataset, including their previous attrition, demographic variables, work experience, employment situation, compensation, organizational support, job attitudes, occupational stress, and work-life balance. The workforce-performance data were gathered from an openly available employee survey that focuses on innovative HR strategies, digital HR technologies, employee readiness, organizational support culture and employee performance (Ali, 2026). A total sample of 368 respondents and 26 scale items on five-point Likert scales were included in the data set.

2.3 Outcome and Predictor Construction

The original attrition response was recoded as a binary outcome, with 1 representing retained employees and 0 representing respondents reporting previous attrition. Several other factors were not included in the primary model as they may be close to the reported attrition outcome, such as health issues, job opportunities, and years of experience in the previous organization. The resulting leakage-reduced feature set included demographic, occupational, compensation, development, support, satisfaction, commitment, stress and work-life characteristics. Five employee-performance items were used to calculate the mean as the outcome of the workforce-performance analysis. These four predictor scores were derived as the means of the items within each subscale: innovative HR strategies, innovative HR technologies, employee readiness and organizational support culture. Five items were used for the constructs, with the exception of organizational support culture, which had six items. All items were retained for the primary analysis based on the items that were coded.

2.4 Data Preparation and Measurement Assessment

Any inconsistencies in the capitalization, spacing and spelling of variable names and categorical responses were removed to standardize the data. Identifiers of the respondents were kept for validation but not for modeling. Categorical levels that were encountered less than 5 times were grouped in pre-processing. One-hot encoding was done for categorical predictors, and no categorical encoding was done for numerical predictors, which were passed to the estimator. To avoid the information from the validation folds affecting the model training, all the steps of preprocessing were handled in the modeling pipelines. The number of missing values, complete duplicate records, duplicate profiles of the predictors, and unique identifiers were used to assess the quality of the datasets. The internal consistency of each scale measuring workforce performance was assessed by calculating Cronbach's alpha and corrected item-total correlations. Distribution of the composite scores was expressed as mean, standard deviation, median, range and interquartile range. Spearman correlation was used to analyze between the five constructs for workforce performance.

2.5 Predictive Modeling and Validation

Two classification methods, logistic regression and random forest, were tested for employee attrition. Three different algorithms were tested for workforce performance: multiple linear regression, random forest regression and a mean regressor. To ensure reproducibility, a random seed of 42 was used with random forests, which had 400 trees. The model comparison was based on repeated (five-fold) cross-validation performed 3 times.

The area under the receiver operating characteristic curve, balanced accuracy, class-specific recall, and Matthews correlation coefficient were used to compare retention models. Three different workforce-performance models were evaluated using R-squared, mean absolute error, and root mean squared error. The best model for each outcome was chosen based on cross-validated predictive ability and the suitability of the model for the analytical goal. Out-of-fold predictions made using five-fold cross-validation were used to estimate final performance. Moreover, a confusion matrix was used to evaluate the retention model at the classification threshold of 0.50. Primary models: 95% confidence intervals were estimated on the basis of 1,000 bootstrap replications of the out-of-fold predictions. The sixth item of the organizational-support scale was reverse-scored to measure the sensitivity of the workforce-performance estimates in this scale to different coding of the items.

2.6 Explainability and Subgroup Assessment

Permutation importance within held-out folds of cross-validation was used to assess the importance of each of the global models. The mean reduction in Validation ROC-AUC and Validation R-squared for workforce performance was used as a measure of permutation importance. The SHAP values were then computed using the fitted random forests to estimate the contribution of each predictor to the model output. First, one-hot-encoded features associated with each predictor were collapsed back to the predictor name and then mean absolute SHAP values were computed. The direction of the contribution of each workforce-performance predictor was determined by the Spearman correlation coefficient between the observed value of the predictor and an associated SHAP value. Out-of-fold predictions were performed in the retention model in order to evaluate the model separately for the female and male subgroups. Subgroup assessment included ROC-AUC, balanced accuracy, predicted retention rate, retention precision, and class-specific recall. Absolute differences between subjects were reported for demographic parity, equal opportunity, false-positive rate, equalized-odds, balanced accuracy, and ROC-AUC differences.

3. Results

3.1 Dataset characteristics and outcome distribution

The employee-retention dataset contained 1,191 observations and 36 analytical variables, while the workforce-performance dataset contained 368 observations and 32 variables. Neither dataset contained missing values or complete duplicate records. One duplicate predictor profile was identified in the retention data, but the records had distinct respondent identifiers and were therefore retained. Table 1 shows the

structural characteristics of the two analytical datasets.

Table 1. Structural characteristics of the analytical datasets

Dataset	Observations	Variables	Missing values	Complete duplicate rows	Duplicate predictor profiles	Unique identifiers
Employee retention	1,191	36	0	0	1	1,191
Workforce performance	368	32	0	0	0	368

Among the 1,191 respondents in the retention dataset, 676 respondents, representing 56.8%, were classified as retained, while 515 respondents, representing 43.2%, reported previous attrition. Neither outcome category represented less than 40% of the sample.

3.2 Workforce-performance scales and descriptive distributions

All five workforce-performance scales demonstrated high internal consistency. Cronbach's alpha ranged from 0.845 for employee performance to 0.901 for organizational support culture. The minimum corrected item-total correlation was 0.627, and the maximum was 0.798. Table 2 shows the reliability estimates for the five multi-item constructs.

Table 2. Reliability of the workforce-performance scales

Construct	Number of items	Cronbach alpha	Minimum corrected item-total correlation	Maximum corrected item-total correlation
Innovative HR strategies	5	0.898	0.691	0.786
Innovative HR technologies	5	0.894	0.712	0.798
Employee readiness	5	0.886	0.646	0.787
Organizational support culture	6	0.901	0.684	0.794
Employee performance	5	0.845	0.627	0.678

The mean composite scores were 3.401 for innovative HR strategies, 3.366 for innovative HR technologies, 3.532 for employee readiness, 3.557 for organizational support culture, and 3.560 for employee performance. Median scores ranged from 3.400 to 3.667. Employee performance had a standard deviation of 0.703 and ranged from 1.200 to 5.000. The strongest bivariate relationship with employee performance was observed for employee readiness, with a Spearman correlation of 0.729. Employee performance was also positively correlated with innovative HR strategies, with a Spearman correlation of 0.605; organizational support culture, with a Spearman correlation of 0.545; and innovative HR technologies, with a Spearman correlation of 0.544. Figure 1A shows that 56.8% of respondents were classified as retained and 43.2% reported previous attrition. Figure 1B shows broadly comparable score distributions across the five workforce-performance constructs, with slightly higher median values for employee readiness, organizational support culture, and employee performance.

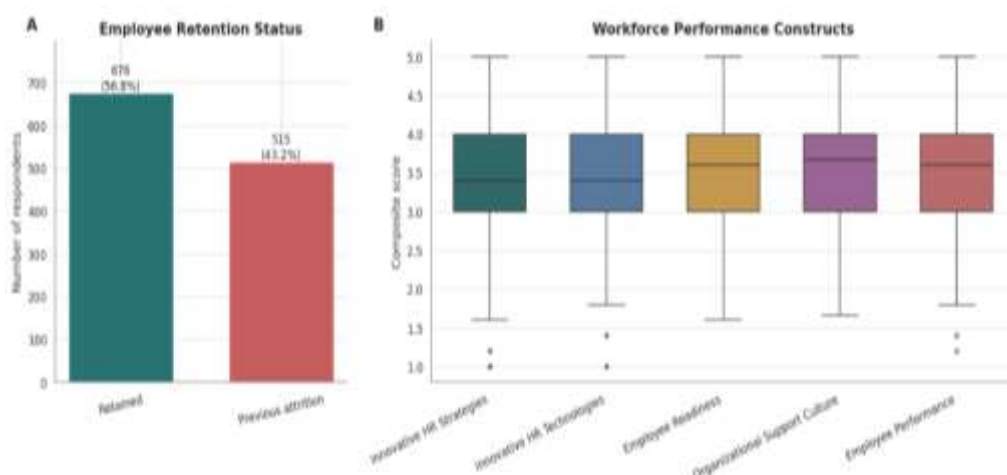


Figure 1. Distribution of the study outcomes and constructs: (A) employee-retention status; (B) workforce-performance construct scores.

3.3 Predictive performance of the primary models

The leakage-reduced random forest was used as the primary retention model. This model excluded health issues, job opportunities, and years of experience in the previous organization. In repeated cross-validation, it produced a mean ROC-AUC of 0.824, balanced accuracy of 0.734, retention recall of 0.747, attrition recall of 0.722, and Matthews correlation coefficient of 0.468. The out-of-fold retention predictions produced an ROC-AUC of 0.824, with a 95% bootstrap confidence interval from 0.801 to 0.847. Balanced accuracy was 0.735, retention recall was 0.744, attrition recall was 0.726, and the Matthews correlation coefficient was 0.468. The model correctly classified 503 of 676 retained respondents and 374 of 515 respondents with previous attrition. It classified 173 retained respondents as previous attrition and 141 respondents with previous attrition as retained. For workforce-performance prediction, the random forest produced a repeated cross-validation mean R-squared of 0.684, compared with 0.554 for multiple linear regression and -0.024 for the mean baseline. The corresponding mean root mean squared error values were 0.387, 0.460, and 0.701, respectively. The random forest's out-of-fold workforce-performance predictions yielded an R-squared of 0.686, with a 95% bootstrap confidence interval from 0.620 to 0.740. The mean absolute error was 0.296, the root mean squared error was 0.393, and the Spearman correlation between observed and predicted scores was 0.817. Table 3 shows the primary out-of-fold results for the retention and workforce-performance models.

Table 3. Out-of-fold predictive performance of the primary models

Analysis	Sample size	Primary model	Primary metric	Estimate	95% confidence interval	Secondary metric	Estimate
Employee-retention classification	1,191	Random forest	ROC-AUC	0.824	0.801–0.847	Balanced accuracy	0.735
Workforce-performance regression	368	Random forest	R-squared	0.686	0.620–0.740	Root mean squared error	0.393

Note: ROC-AUC refers to the area under the receiver operating characteristic curve. Estimates were calculated from out-of-fold predictions.

The leakage-reduced employee-retention model has an out-of-fold receiver operating characteristic curve, as shown in Figure 2A. The relationship between out-of-fold predicted and observed workforce-performance scores is shown in Figure 2B.

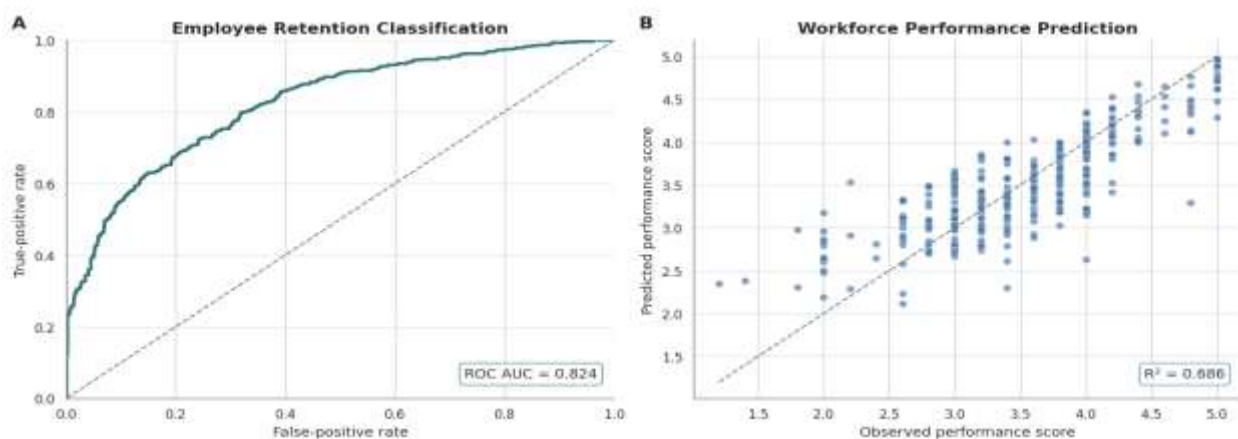


Figure 2. Out-of-fold predictive-model validation: (A) receiver operating characteristic curve for employee-retention classification; (B) observed and predicted workforce-performance scores.

3.4 Global model explanations

Years of experience had the highest mean reduction in the validation ROC-AUC of 0.063 for cross-validated permutation importance. Job stability, training participation in the last 3 years, job title, promoted deservedly, and medical insurance had the next highest permutation importance values. Amongst the predictors of retention, years of experience were ranked first and were the most important on the aggregated SHAP analysis. It had the highest mean absolute SHAP value of 0.063; the other variables were: job stability 0.047, deserved promotion 0.035, emotional commitment 0.029, and participation in training programmes 0.027. The permutation importance for workforce performance was the highest, with a mean reduction in the validation R-squared of 0.561, indicating that employee readiness was the most important factor. After organizational support culture had a score of 0.299, innovative HR strategies scored 0.122 and innovative HR technologies scored 0.053. The same order was obtained with the SHAP ranking. The global SHAP values for the two models are shown in Table 4.

Table 4. Global SHAP importance of employee-retention and workforce-performance predictors

Analysis	Rank	Predictor	Mean absolute SHAP value
Employee retention	1	Years experience	0.063
Employee retention	2	Job stability	0.047
Employee retention	3	Deserved promotion	0.035
Employee retention	4	Emotional commitment	0.029
Employee retention	5	Training programmes during the previous three years	0.027
Employee retention	6	Rewards and wages satisfaction	0.025
Employee retention	7	Sector	0.020
Employee retention	8	Medical insurance	0.018
Employee retention	9	Job title	0.017
Employee retention	10	Department	0.016
Workforce performance	1	Employee readiness	0.328
Workforce performance	2	Organizational support culture	0.147
Workforce performance	3	Innovative HR strategies	0.119
Workforce performance	4	Innovative HR technologies	0.059

Note: SHAP refers to Shapley additive explanations. SHAP magnitudes are comparable within each model but not directly between the classification and regression models.

The top 10 retention predictors by the aggregated mean absolute SHAP values are displayed in Figure 3A. The SHAP importance of the four workforce-performance predictors is displayed in Figure 3B.

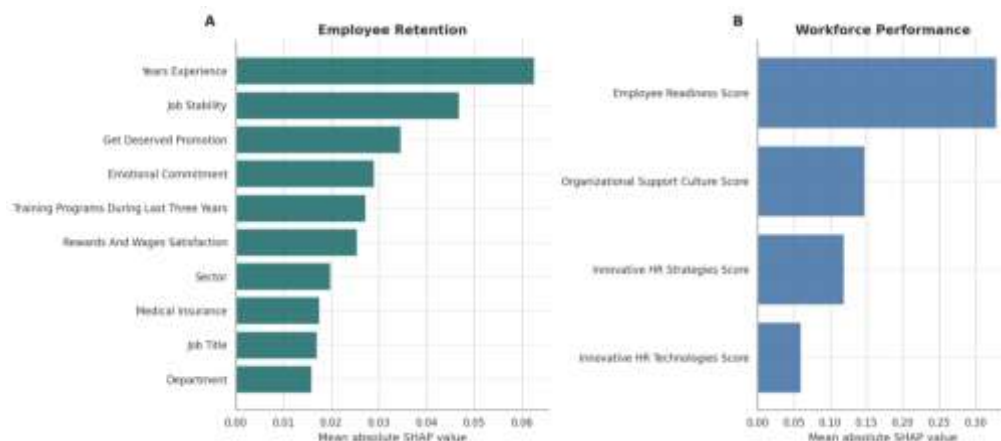


Figure 3. Global explainability of the primary random forest models: (A) SHAP importance of employee-retention predictors; (B) SHAP importance of workforce-performance predictors.

3.5 Directional explanations and subgroup model performance

All four workforce-performance predictors had positive correlations between their observed values and SHAP contributions. The Spearman correlation for employee readiness was the highest (0.981), all with a positive direction. It was followed by innovative HR strategies (Spearman correlation of 0.829), innovative HR technologies (Spearman correlation of 0.756) and organizational support culture (Spearman correlation of 0.603). The out-of-fold R-squared, mean absolute error, root mean squared error, and Spearman correlation of the sensitivity analysis where the sixth organizational support item was kept as recorded were 0.686, 0.296, 0.393 and 0.817, respectively. The R-squared dropped to 0.673, and the two errors rose to 0.306 and 0.401, respectively, when the item was reverse-scored. The Spearman correlation went down a bit to 0.814. The gender-subgroup analysis was based upon 678 females and 513 males. The observed retention rate was 0.608 with a predicted retention rate of 0.593 for female respondents. The figures were 0.515 and 0.472, respectively, in the case of the male respondents. The AUC for females and males were 0.812 and 0.833, respectively. For both, the balanced accuracy was 0.722 and 0.744. The results from the models are presented in Table 5 separately by gender.

Table 5. Gender-subgroup performance of the employee-retention model

Subgroup	Sample size	Observed retention rate	Predicted retention rate	Mean predicted retention probability	ROC-AUC	Balance accuracy	Retention precision	Retention recall	Attrition recall
Female	678	0.608	0.593	0.563	0.812	0.722	0.786	0.767	0.677

Male	513	0.515	0.472	0.512	0.833	0.744	0.773	0.708	0.779
------	-----	-------	-------	-------	-------	-------	-------	-------	-------

The positive direction of the correlations between the workforce-performance predictors and their SHAP contributions is depicted in Figure 4A. The observed, predicted retention rates, class-specific recall and balanced accuracy are shown for the female and male subgroups in Figure 4B.

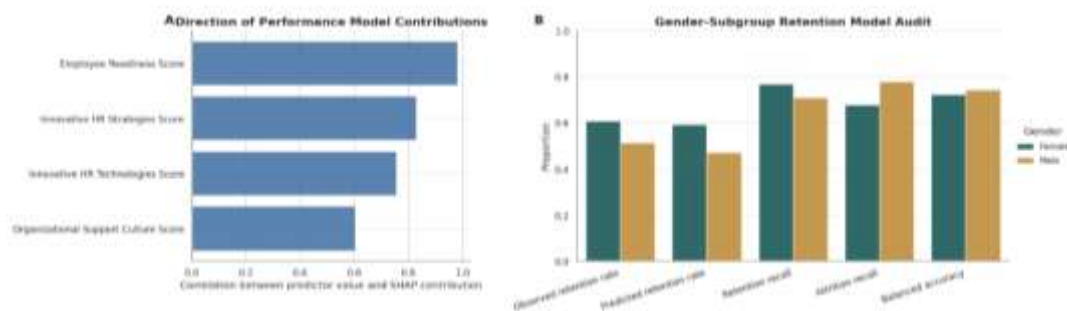


Figure 4. Directional explainability and retention-model fairness audit: (A) direction of workforce-performance model contributions; (B) gender-subgroup assessment of the employee-retention model.

3.6 Between-group fairness gaps

The difference between the predicted retention rates of the gender subgroups was 0.121, an absolute number. The difference in the equal opportunities recall (retained) was 0.059. The difference between the false positive rates was 0.102, which led to an equalized-odds difference of 0.102. The differences in overall predictive performance were smaller: 0.022 for the balanced-accuracy difference, and 0.021 for the ROC-AUC difference. The between-group fairness gaps calculated are presented in Table 6.

Table 6. Between-group fairness gaps for the employee-retention model

Fairness measure	Absolute gap	Operational definition
Demographic parity difference	0.121	Difference in predicted retention rates
Equal opportunity difference	0.059	Difference in correctly identified retained employees
False-positive-rate difference	0.102	Difference in previous-attrition cases predicted as retained
Equalised-odds difference	0.102	Maximum difference across retention recall and false-positive rate
Balanced-accuracy difference	0.022	Difference in subgroup balanced accuracy
ROC-AUC difference	0.021	Difference in subgroup discrimination

Note: Fairness gaps were calculated from out-of-fold predictions using a classification threshold of 0.50.

4. Discussion

The results show that explainable machine learning can be beneficial in predicting employee turnover and workforce performance, and also maintain a reasonable level of interpretability for an organization. The leakage-reduced random forest correctly classified the retained employees and the respondents who reported previous attrition, with a score of 0.824 on ROC-AUC and 0.735 on balanced accuracy. The distribution of the two outcomes was fairly equal for classification, with no perfect matchings for either of the outcomes, retention or attrition. The model should be viewed, then, as a tool to help in the identification of a pattern of heightened risk but not as a way to predict what an individual employee will do in the future. The most important SHAP values and permutation importance were years of experience. Other factors such as job security, appropriate promotion opportunities, emotional investment, engagement in recent training, and level of satisfaction with rewards and wages also had some influence. The pattern suggests the retention predictions were based on a combination of career experience and employees' beliefs regarding continuity, development, recognition, and attachment. The agreement between the two explanation methods provides support for the validity of the rankings and the fact that they were not due to one specific methodology in the interpretation process. The workforce-performance model explained 68.6% of the out-of-fold variation in employee-performance scores, and was superior to multiple linear regression and the mean baseline. This advantage was indicative of a relationship between performance and nonlinearities or interactions, picked up better by the random forest. The most prevalent factor was employee readiness, followed by organizational support culture, innovative HR strategies, and digital HR technologies. Construct values correlated with SHAP values were positive, indicating that the higher the score on the construct, the more the construct contributed to an increase in the predicted level of performance. The relatively minor part of digital HR technologies suggests that technology is not as informative as employees' willingness to use change productively and the assistance they receive from the organization.

The retention results are consistent with research using XAI that interpretable models can facilitate the generation of explanations for managerial decisions, alongside providing useful discrimination, at the level of the predictors (Baydili & Tasci, 2025). The usefulness of their approach is also backed by research on the way predictive HR analytics can help plan for the retention of talent, as they can detect patterns in the workforce before the standard HR review processes (Căvescu & Popescu, 2025). The learning points of how the random forest model outperforms the linear models align with the other applications of machine learning models to convert multidimensional workforce data into actionable HR insights (Saha et al., 2020). Similarly, a range of studies on deep learning have shown that sophisticated predictive techniques can enhance the analysis of retention when there are non-linear relationships between employee attributes (Sharma et al., 2024). These high scores of employee readiness and organizational support further align with research showing that AI is a tool to assist in workforce forecasting, not supplanting organizational judgement (Joseph et al., 2024). The critical factors of training, promotion, stability, and rewards (Paigude et al., 2023) are also substantiated by evidence that AI can strengthen retention through the early recognition of employee needs. The present explanation results expand analytics-based retention efforts by not only presenting those employees that might have lower predicted probabilities, but also identifying features that lead to these outputs (Sendil et al., 2024). This visionary approach complements intelligent systems for retention, where attrition predictions are linked to targeted HR measures (Lenka et al., 2025). The results also support the proposed usage of AI in talent management and retention and indicate that human-centric factors are still important in predicting the conditions (Saadallah, 2025). In addition to the development of AI and survival analysis methods that focus on robust estimation of employee-retention outcomes (Folorunso et al., 2025), the use of validated out-of-fold estimates adds to the existing methods. The added value of subgroup results is that while both female and male employees had similar ROC-AUC and balanced accuracy, they differed in their predicted retention rates and errors of class. The model, therefore, showed the same level of overall discrimination for both genders, but needed to focus on the distribution of errors.

HR managers can leverage the factors identified to focus on career development, viable promotion systems, job security, training opportunities, commitment of staff, and fair reward systems. Employee readiness and organizational support should be part of performance initiatives, along with investing in digital HR tools. SHAP explanations can support managerial interpretation, but predictions should be considered alongside organizational evidence and employee context. Integrated subgroup monitoring during deployment should continue because even if aggregate accuracy is achieved, there is no guarantee that the pattern of errors is the same.

There are a few limitations one should be aware of. There is a possibility of differences in findings within organisations, industries and countries. The models need to be validated by longitudinal and organization-specific data, future changes in the risk of retention need to be explored, and research into the effectiveness of readiness-focused interventions and support-focused interventions must be conducted to determine their impact on performance. These models would be further validated by external validation and continued fairness auditing for their use in operational HR settings.

5. Conclusion

Explainable Artificial Intelligence (XAI) offered a clear tool for forecasting employee turnover and workforce performance, and helped to illuminate the reasons behind model results. The random forest was useful for discriminating between workers who remained with the organization and those who left and accounted for a significant amount of variance in the performance of the workforce. Years of experience, job stability, opportunities for promotion, emotional commitment, training participation and satisfaction with rewards and wages were the most important factors in retention predictions. The most dominant predictor of workforce performance was employee readiness, followed by organizational support culture, innovative HR and digital HR technologies. The results suggest that the best time for technological investment is when workers are open to change and have support from their organization. The SHAP analysis enhanced the transparency of the model by individually highlighting the contribution of the workforce characteristics to the prediction, whereas the gender assessment highlighted the same overall predictive ability but different error patterns within the subgroups. The models need to be used to supplement, not replace, sound HR decision-making. These insights can be used by organisations to improve employee development, job security, employee recognition, organisational support and responsible workforce planning. Validation and fair monitoring are still needed before going into operation.

References

1. Abonamah, A., La Torre, D., Poulin, M., & Repetto, M. (2022, October). Explainable artificial intelligence in human resources: a computational study. In 2022 International Conference on Data Analytics for Business and Industry (ICDABI) (pp. 1-6). IEEE.
2. Ali, R. H. (2026). Innovative HR strategies and digital HR technologies driving employee performance in digital transformation: The mediating role of employee readiness and the moderating role of organizational support culture (Version 1) [Data set]. Mendeley Data. <https://doi.org/10.17632/xbg5gzskfv.1>

3. Alqahtani, H., Almagrabi, H., & Alharbi, A. (2025). Saudi employee attrition dataset (Version 1) [Data set]. Mendeley Data. <https://doi.org/10.17632/6z2hty8php.1>
4. Azeez, S. A. (2017). Human resource management practices and employee retention: A review of literature. *Journal of Economics, Management and Trade*, 18(2), 1-10.
5. Baydili, İ. T., & Tasci, B. (2025). Predicting employee attrition: XAI-powered models for managerial decision-making. *Systems*, 13(7), 583.
6. Bos-Nehles, A. C., & Veenendaal, A. A. (2019). Perceptions of HR practices and innovative work behavior: the moderating effect of an innovative climate. *The International Journal of Human Resource Management*, 30(18), 2661-2683.
7. Bos-Nehles, A., Renkema, M., & Janssen, M. (2017). HRM and innovative work behaviour: A systematic literature review. *Personnel review*, 46(7), 1228-1253.
8. Căvescu, A. M., & Popescu, N. (2025). Predictive analytics in human resources management: Evaluating AIHR's role in talent retention. *AppliedMath*, 5(3), 99.
9. DiRomualdo, A., El-Khoury, D., & Girimonte, F. (2018). HR in the digital age: How digital technology will change HR's organization structure, processes and roles. *Strategic HR Review*, 17(5), 234-242.
10. Folorunso, S. A., Kehinde, R. O., & Folorunso, M. A. (2025). Predicting employee retention using artificial intelligence and survival analysis approaches. *Journal of Artificial Intelligence in Engineering Practice*, 2(1), 31-38.
11. Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence. *Science robotics*, 4(37), eaay7120.
12. Hom, P. W., Allen, D. G., & Griffeth, R. W. (2019). *Employee retention and turnover: Why employees stay or leave*. Routledge.
13. Joseph, A., Johnson, A., & Cheriyan, A. (2024). Workforce Forecasting Using Artificial Intelligence. *Artificial Intelligence in Forecasting*, 146-172.
14. Lenka, R., Chitranshi, J., & Pabalkar, V. (2025, July). An Artificially Intelligent System to Strategize and Predict Employee Attrition and Retention in HR Management. In *International Conference on ICT for Sustainable Development* (pp. 213-223). Cham: Springer Nature Switzerland.
15. Paigude, S., Pangarkar, S. C., Hundekari, S., Mali, M., Wanjale, K., & Dongre, Y. (2023). Potential of artificial intelligence in boosting employee retention in the human resource industry. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(3s), 01-10.
16. Pradhan, R. K., & Jena, L. K. (2017). Employee performance at workplace: Conceptual model and empirical validation. *Business perspectives and research*, 5(1), 69-85.
17. Ramachandran, K. K., Mary, A. A. S., Hawladar, S., Asokk, D., Bhaskar, B., & Pitroda, J. R. (2022). Machine learning and role of artificial intelligence in optimizing work performance and employee behavior. *Materials Today: Proceedings*, 51, 2327-2331.
18. Saadallah, F. O. (2025). The Impact of Artificial Intelligence on Talent Management and Employee Retention. *Rihan Journal for Science Publishing*, (59), 325.
19. Saha, B., Chhapola, A., & Jain, A. (2020). AI-driven workforce analytics: Transforming HR practices using machine learning models. *International Journal of Research and Analytical Reviews*, 7(2), 982-997.
20. Schneider, B., Yost, A. B., Kropp, A., Kind, C., & Lam, H. (2018). Workforce engagement: What it is, what drives it, and why it matters for organizational performance. *Journal of Organizational Behavior*, 39(4), 462-480.
21. Sendil, M. S., Kannan, G., Fufa, G., Shyamala, L., Hegde, N. P., & Meenakshi, S. (2024, November). Data Analytics and Artificial Intelligence for Improving Employee Retention Strategies. In *2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)* (Vol. 1, pp. 250-254). IEEE.
22. Sharma, R., Jain, A., & Manwal, M. (2024, June). Enhancing human resource management through deep learning: A predictive analytics approach to employee retention success. In *2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)* (pp. 1-4). IEEE.
23. Singh, D. (2019). A literature review on employee retention with focus on recent trends. *IJSRST*.
24. Zhang, J., & Chen, Z. (2024). Exploring human resource management digital transformation in the digital age. *Journal of the knowledge economy*, 15(1), 1482-1498.