Volume 12, No.8, August 2025
Journal of Global Research in Mathematical Archives
RESEARCH PAPER
Available online at http://www.jgrma.com
ISSN 2320 – 5822

PERFORMANCE ANALYSIS OF MACHINE LEARNING BASED CLASSIFIERS FOR PREDICTING EMPLOYEE ATTRITION IN RESOURCE-CONSTRAINED HR SYSTEMS

Dr. Nilesh Jain[1]

[1] Associate Professor, Department of Computer Science and Applications, Mandsaur University, Mandsaur, India, nileshjainmca@gmail.com

Abstract: The problem of employee turnover is getting worse for many tech companies throughout the world. Staff turnover, sometimes known as attrition, is on the rise and particularly high in the information technology sector for reasons that are difficult to predict or predictably address. Any industry, no matter how big or little, can benefit from this. Keeping an existing employee costs much less than recruiting a new one. Acquiring a new resource necessitates a period of adjustment to a new team's or company's culture as well as training and the transfer of existing knowledge. Using the IBM HR dataset obtained from Kaggle, this project endeavors to forecast employee turnover in HR systems with limited resources. Data cleaning, encoding, normalization, feature extraction, and class balancing using SMOTE were all steps in the extensive pretreatment pipeline that was applied to the dataset, which contains 1,470 employee records with 35 features. Evaluate the predictive power of various ML models, including LR, RF, SVM, and a proposed DNN. The DNN model achieved the best results of all the models when it came to capturing complex patterns in the data. It achieved an F1-score of 94.52% and had very high accuracy and precision. Using deep learning approaches to enhance HR decision-making in environments with limited resources is essential for employee retention.

Keywords: Employee Attrition, Human Resource, Machine Learning, Attrition Rate, Organization, Analysis, Employee Attrition Causes.

1. INTRODUCTION

The competition among corporations and businesses is highly dependent on the throughput of the employee. The key to contributions of steady and cooperative employees are constructing and maintaining the relevant atmosphere. Staff turnover has been a sensitive issue in an organization especially in a resource-stretched setting where human capital has been a big determinant of productivity and efficiency of operations [1][2]. Understanding the practices of the workforce and its relation to stability and throughput of the enterprise, since the establishment of the beneficial atmosphere within an organization is one of the priorities of the activities of Human Resource (HR) departments [3][4]. This is important as it helps in examining the employee history and behavior pattern so as to create a conducive environment and reduce attrition. Attrition of employees is a process where employees move away or out of an organization as a result of many factors including poor pay, job dissatisfaction, too much work etc or by finding greener pastures outside the organization [5]. Bad turnover rates cause the loss of trained and experienced members which translates to low productivity, high employment expenses and business continuity [6][7][8]. Inflows, on the other hand, ensure that there is continuity of knowledge in the organization through retention of employees and therefore ensure that less is spent in the hiring process.

Attrition occurs in various forms: voluntary, when employees because of some reasons decide to leave an organization, involuntary, when an organization fires employees, external, when employees decide to leave an organization and join another organization and internal, when the employees decide to change positions within an organization [9][10]. The calculation of the attrition rate, which is the ratio of the number of employees who would have left to the mean workforce is used to gain very important information about workforce stability and organization health.

The recent years along with the growth of HR analytics have seen the popularization of machine learning (ML) methods in the prediction and management of employee attrition. These models utilize past workforce statistics to come up with predictive data to facilitate robust decision-making and planning of resources [11][12]. Introduction of ML-based forecasting analytics can give the necessary impetus to organizations to predict the high-risk employees, thereby reworking the retention strategies. Yet, there is also the problem of interpretability and transparency of the traditional ML models when it comes to a more complex and dynamic attrition case [13][14][15]. Since the validity and interpretability of prediction results are essential aspects of any HR system, the paper offers a comparison between different machine learning classifiers in predicting worker arteriosclerosis with special interest in their functionality within limited HR contexts. The aim is to assess the performance, scaling, and interpretability of these models to shortlist practical models of deployments in the real world.

1.1. Motivation and Contribution

This research is motivated by the increasing difficulty experienced by firms in maintaining talent especially in financially limited human resource conditions whereby high levels of employee turnover may result in profound operational and financial losses. Conventional ways of detecting attrition hazard usually make use of manual reviewing or unsophisticated models, thus missing the inclusiveness of employee behavior that is complex and non-linear. As the workforce data becomes more readily accessible, and machine learning becomes more advanced, it becomes highly incentivized to build smart predictive systems that able to identify the employees about to quit before they do it. With the help of the IBM HR dataset and the use of more advanced methods, including deep learning, the study give power to HR departments, providing them with evidence-based data to make more timely and rational decisions on the retention process, which lead to the enhanced stability of the workforce and organizational efficiency. There are three important contributions about the present research to Resource-Constrained HR Systems:

- Utilized the IBM HR dataset from Kaggle, comprising 1,470 records and 35 features, ensuring real-world applicability.
- Understood data preprocessing comprehensively, including missing values, noise and outliers, and did it to enhance the data quality.
- Performed normalization of the feature distributions using Applied Standard Scaler to train the model in the best way.
- Balanced the data by imposing SMOTE (Synthetic Minority Over-sampling Technique) in order to increase fairness and the performance of the model.
- Created an effective model of predicting employee attrition with Deep Neural Network (DNN) architecture optimized to operate within a resource-scarce HR setting.
- A wide range of metrics were employed to assess the model's efficacy, including precision, accuracy, recall, and F1-score, to guarantee its thoroughness.

1.2. Justification and Novelty

The rationale behind the study is that adequate and largescale employee attrition prediction models are crucial yet the traditional methods remain ineffective in resource-limited HR systems, whose analytical resources and nuances of workforce data are limited. The novelty aspect of this research is the deployment of such (DNN) model that outperforms the classical models (such as, LR, RF, and SVM) as, it effectively represents non-linear relationships and high-dimensional interaction effects in the employee data. Along with that, SMOTE as a class balancing technique, feature extraction method and standardized preprocessing also contribute to the robustness and generalizability of the model. Such a combination is a most useful and new approach to modern HR analytics, as it shows much higher accuracy and reliability of prediction.

1.3. Structure of the Paper

The structure of the paper is as follows: Section 1 gives the introduction of a problem of employee attrition in resource-constrained HR systems and the role of predictive modeling. Part 2 conducts review of related literature on machine learning methods to use in prediction of attrition. The methodology of the research, i.e. the description of a dataset, preprocessing, and applying of classifiers, is described in Section 3. Section 4 shows and makes the discussion involving performance comparison of models, with DNN being the most accurate. In Section 5, the study is closed, and future work based on the methods of explainable AI, hybrid models, and time-sequence forecasting are offered.

2. LITERATURE REVIEW

This study has reviewed and analyzed a broad base of important research studies on the ability to predict employee attrition within resource-constrained HR systems to inform and reinforce the process of its design within the present study.

Tanmayi et al. (2025) in the suggested project the ML, Ensemble models and DL are applied to forecast the employee attrition in such a way that it assists the firm to shoot their target to an extent where they could please their employees. These models are trained and tested based on one of the datasets available on Kaggle. Most of the models such as CatBoost, AdaBoost, Random Forest, SVM, DNN, Ensemble model of CatBoost and AdaBoost, and Ensemble model of hard Voting Classifier are applied. The above results have been generated by random forest giving the highest value compared to the other models used, which were 93.1 percent as a f-score and 93.5 % accuracy [16].

M et al. (2025) presents a new approach to machine learning-supported employee categorization to solve one of the most burning problems of employee turnover and introducing an outstanding contribution to the sphere of employee management. Based on a dataset that belongs to XYZ Corporation, the system applies advanced machine learning abilities to model the extremely accurate and interpretable staff turnover rates. The best algorithm turned out to be the XGBoost model that reached an accuracy of 92.3%, a precision of 89.7%, and an F1-score of 90.6%. The SHAP (SHapley Additive Explanations) values were used to identify key predictors of attrition, including job satisfaction, work-life balance, and compensation, allowing the HR professionals to get an insight into the underlying causes of turnover and being able to solve it. A key added value to the study is to make such fairness-informed approaches the principle of curbing bias in employee classification, such as re-weighting and adversarial debiasing [17].

Muthugala et al. (2024) uses machine learning (ML) as a predictor of employee retention and attrition in the recruitment segment. Projections can help people in charge make the most adequate choice of long-term employees and minimize the risk of losing the working staff on the fly. The analysis was based on a corporate human resource data of a medium scale company, that had both pre- and post-recruitment data. The author has selected the algorithms of SVM, Decision Tree, and Random Forest to build a model according to the literature. Relevant metrics from six experiments using different inputs were presented. The Random Forest (RF) classifier outperformed the others in five of the six experiments. The RF algorithm outperformed both models, with an accuracy of 94.5% for retention and 89.7% for attrition prediction [18].

Ozakca, Bulus and Cetin (2024) was made available as SaaS (Software as a Service) based in order to facilitate accessibility to the target audience. For the analysis phase, the columns in the data set were analyzed with Pandas and Scikit-Learn library. In order to enrich the data set and increase the efficiency of the model, artificial intelligence-based synthetic data generation (Data Augmentation) was performed. In case of missing columns in the trained model, KNN-Data Imputation, one of the missing data completion methods, was used. In the optimization process of the model, hyperparameter optimizations were performed to achieve maximum efficiency with improvements. 5-fold cross-validation prevented the model from over-learning. Performance was analyzed on the basis of Accuracy, Recall, Precision and F1-score metrics and the success criterion was determined as F1-Score. The model was presented as a web service with an accuracy rate of 0.99% [19].

Jin et al. (2024) aims to analyze such factors and predict employee attrition using ML and ensemble learning methods. A comparison of four ml methods and three ensemble learning methods was conducted, utilizing grid search for hyperparameter tuning to identify the optimal model. Also, exploratory data analysis was used to analyze employee data in order to determine the factors that would lead to employee attrition. The findings suggest that RF has the greatest level of accuracy which is nearly 95%. Some of the main factors that cause employee attrition are overtime, stock options level and satisfaction of the job. The findings can help organizations to curb employee turnover by eliminating the major factors of turnover and they can increase job satisfaction and help the organizations to gain better stability and image in the marketplace [20].

Silpa et al. (2023) Enriched Employee Retention Analysis System (EERAS) is a new approach that uses the correct combination of machine learning techniques and feature selection paradigms. In order to keep employees from leaving, the suggested EERAS has two goals: first, to identify the causes of employee turnover, and second, to propose solutions. The authors achieve their goals by conducting an empirical study on a large employee dataset using four feature selection methods: MRMR Chi 2, ANOVA, and Kruskal. The study compares five different classification learning algorithms: Decision Tree, Linear Regression, Naive Bayes, Support Vector Machine, and XGBoost. Results showed that feature selection improves classification accuracy and shortens training time. To further aid in the implementation of HR policy, the system also provides recommendations on critical factors that can guarantee the continued employment of judicial staff [21].

Maharana et al. (2022) focus on amassing an enormous quantity of data and accommodating ever-evolving preferences. In light of this, artificial intelligence and machine learning bring a wealth of new possibilities to drive business growth through actionable insights. The study's overarching goal is to provide light on employee turnover intentions, potential deterrents, and ways to use this information to predict future staff turnover. In order to predict employee turnover, this study utilises three machine learning models with the 35-feature IBM Watson dataset. Logistic Regression was the most effective machine learning strategy for this dataset, with an accuracy of 87% and a recall rate of just 0.36 percent, according to the performance metrics [22].

Joseph et al. (2021) purpose of this research is to make predictions about employee turnover and morale in the workplace. The moment an attrition-related survey was created, the data needed to analyze it could be determined by an IT model, which could then anticipate attrition and, based on that, give depression analysis. After applying the preprocessing stages to this dataset, the algorithms utilized to predict the attrition rate were Decision Tree Classifier (DTC), Support Vector Machine (SVM), and Random Forest Classifier (RFC). The resulting accuracy was 86.0 percent. They have anticipated the results using the most important metrics for classification, such as the F1-score and the accuracy metric [23].

Table 1 provides an overview of the recent research on employee attrition prediction based on machine learning including innovative models and datasets used, significant findings, and research limitations.

Table 1: Recent Research on Predictive Modeling of Employee Attrition Prediction by Machine learning

Author(s)	Methodology	Dataset	Key Findings	Challenges/Future Work
Tanmayi et al. (2025)	CatBoost, AdaBoost, Random Forest, SVM, DNN, Voting Ensembles	Kaggle dataset	With an F1-score of 93.1% and an accuracy of 93.5%, Random Forest produced the most favorable results.	Focus on resource-efficiency in deploying ensemble models in low-capacity HR systems
M et al. (2025)	XGBoost with SHAP explainability and fairness-aware techniques	Dataset from XYZ Corporation	Achieved 92.3% accuracy, 89.7% precision, and 90.6% F1-score; identified key features like job satisfaction and compensation	Integration of fairness-aware ML techniques; requires scalability validation
Muthugala et al. (2024)	SVM, Decision Tree, Random Forest	Apparel company's HR dataset	Random Forest performed best with 94.5% accuracy (retention) and 89.7% (attrition)	Further generalization needed beyond sector-specific datasets
Ozakca, Bulus and Cetin (2024)	SaaS ML pipeline with data augmentation, KNN imputation, hyperparameter tuning	Not specified	Model reached 99% accuracy; presented as a web service with 5-fold validation	Generalizability across diverse organizational structures and real-world deployment
Jin et al. (2024)	4 ML + 3 ensemble models with grid search optimization	Employee dataset (unspecified)	Random Forest achieved ~95% accuracy; key factors identified include overtime and job satisfaction	Potential for incorporating temporal dynamics and employee lifecycle modeling

Silpa et al. (2023)	EERAS: 4 feature selection techniques + 5 classifiers	Large employee dataset	Feature selection significantly improved accuracy; recommended retention policies	Need for real-time, adaptive models and integration with HR systems
Maharana et al. (2022)	Logistic Regression, ML on IBM Watson dataset	IBM Watson (35 features)	Logistic Regression achieved 87% accuracy; best recall (0.36)	Improve recall for minority attrition class; explore deep learning approaches
Joseph et al. (2021)	SVM, DTC, RFC + emotional analysis from survey	Custom survey dataset	Achieved 86% accuracy using classification metrics	Expand dataset, include sentiment/emotional analytics, and hybrid models

3. RESEARCH METHODOLOGY

In this study, methodology for predicting employee attrition in resource-constrained HR systems using a Deep Neural Network (DNN) model. The methodology starts by downloading IBM HR Analytics dataset on Kaggle and continues with preprocessing the data by removing the missing values, noise, and outliers. Capital variables were encoded, and the normalization features and Standard Scaler and Feature extraction methods were used. In order to correct the problem of class imbalance, SMOTE was used. The dataset was later divided into training and test datasets. Then a model of DNN was developed and trained with the processed data to extract complicated non-linear patterns. Accuracy, precision, recall, and F1-score are some of the common classification metrics used to assess the model's performance. Figure 1 shows all the steps of the process.

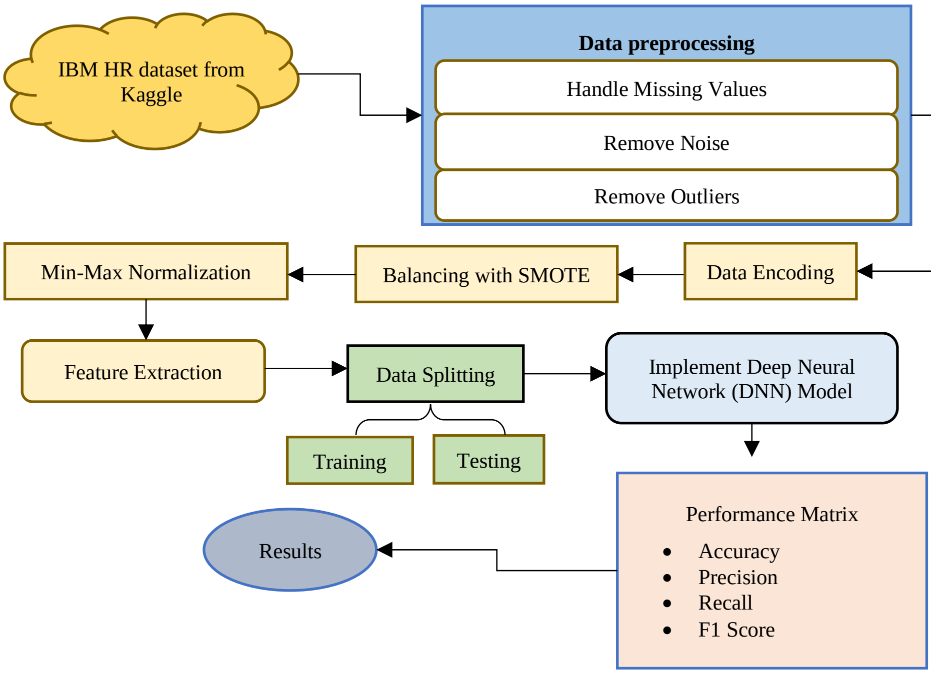

Figure 1: Proposed flowchart for Employee Attrition in HR Systems

Below is a comprehensive explanation of each step in the proposed flowchart for employee attrition prediction within resource-constrained HR systems.

3.1. Data Collection

The IBM HR dataset is a curated collection of employee information that encompasses several attributes relevant to HR analytics. Kaggle Repository was the source of the dataset. A total of 1,470 employees filled out the 35 variables. Data visualizations such as bar plots and correlation matrix were used to examine Attrition distribution, feature correlations etc., and are given below:

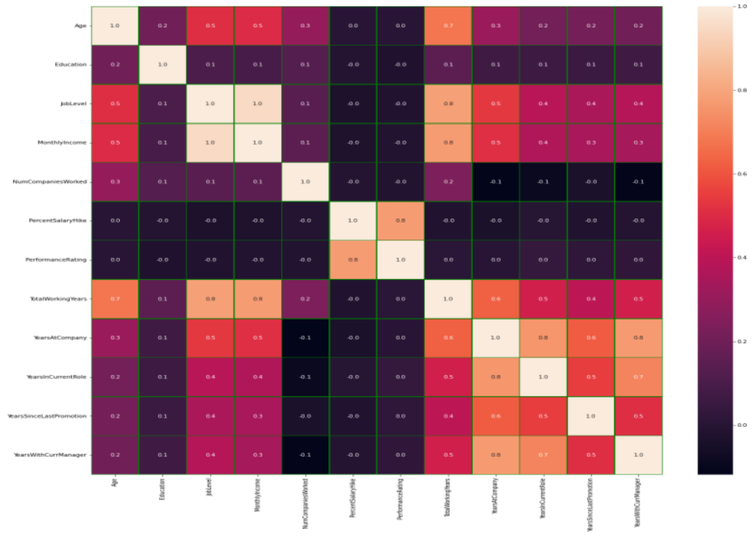

Figure 2: The Dataset Correlation Analysis

Figure 2, a heatmap showing the dataset's correlation matrix for numerical attributes with values between 0 and 1. Darker shades indicate weaker correlations, while lighter shades represent stronger positive correlations. Diagonal values are all 1, indicating perfect self-correlation. Notable correlations include strong relationships between JobLevel and Monthly Income, Total Working Years and Years At Company, and between Years With Curr-Manager and YearsAtCompany, all with coefficients above 0.7. The matrix helps in identifying multicollinearity and feature relationships, which are essential for model building and selection in attrition prediction analysis.

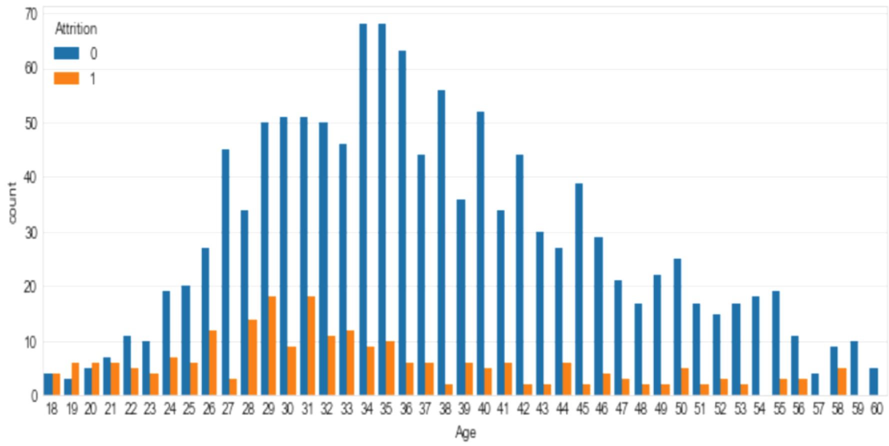

Figure 3: Data Exploration And Analysis

Figure 3 distribution of employee attrition across different age groups is shown in the bar chart. The y-axis shows the total number of employees, and the x-axis shows the ages of those employees, measuring from 18 to 60. Blue bars represent employees who stayed (Attrition = 0), and orange bars represent those who left (Attrition = 1). The chart reveals that attrition is more frequent among younger employees, particularly between ages 26 and 35, with the peak around age 31. As age increases beyond the mid-30s, both overall employee count and attrition rates decline significantly.

3.2. Data Pre-Processing

Data preparation began with collecting the IBM HR dataset from Kaggle, followed by combining and cleaning the data to ensure consistency. Pre-processing of the dataset has been performed to handle missing values, removal of noise and outliers using relevant features that have been extracted. Data transformation and normalization performed between the data transformation and normalization. The most important procedures in the process of preprocessing are as follows:

- **Handle missing values:** Data preprocessing isn't complete without dealing with missing values. The most common strategies are: deletion, imputation (substitution of missing values with some estimation values), using algorithms that are robust to missing data. Data, the reason for missing data, and analysis all play a role in determining the best approach.
- **Remove noise:** De-noising data is most necessary for the good analysis and modeling of it. It applies to this identify and ameliorate unwanted data points or noise that causes important patterns.
- **Remove Outliers:** Data processing includes an outlier removal stage that involves identifying and removing data points that do not fit the pattern with the rest of the data. This is to avoid such extreme values skewing the statistical analysis and machine learning models.

3.3. Data Encoding

Data encoding refers to the process of representing data in a particular format to be used in diverse functionalities such as storage, communication and processing. It is the conversion of data in one form to another that is often numerical or binary representation, in a form more likely to be processed by a given system or application.

3.4. Standardscaler for Data Normalization

A normal distribution with a mean of 0 and a standard deviation of 1 was produced by standardizing the data using the Standard Scaler() technique, which took into account the different scales of each descriptor. This is performed in respect to the mean value of each of the observations and by division of the standard deviation as expressed in Equation (1):

$$z = \frac{x-\mu}{\sigma} \tag{1}$$

with z being the feature's modified value, x being the original value of each descriptor, mu being the feature's mean, and sigma being its standard deviation on the dataset.

3.5. Feature Extraction

Feature extraction has a major impact on the actual results of the model and is hence regarded as the most important concept in deep learning. It can easily teach simulation using this collection of characteristics, and it will significantly improve simulation performance. The goal of feature extraction is to make machine learning models more efficient by reducing large amounts of raw data to a manageable set of features. It's a dimensionality reduction technique that focuses on selecting or creating the most relevant attributes from a dataset, while discarding less important ones. This simplification improves model performance, reduces computational cost, and enhances interpretability.

Figure 4: The pair plot data distributions analysis

Figure 4 pair plot visualizing relationships between various numerical features such as MonthlyIncome, JobLevel, TotalWorkingYears, YearsAtCompany, and YearsWithCurrManager about employee attrition. Each subplot shows scatter distributions and density plots, with blue points representing employees who stayed (Attrition = 0) and red points representing those who left (Attrition = 1). The plot reveals some distinct patterns, such as lower attrition among employees with higher monthly income and job level, and higher attrition rates in early career stages or with fewer years at the company or under the current manager. This visualization aids in identifying correlations and trends relevant to attrition prediction.

3.6. Data Balancing using SMOTE

Data balancing is a method employed in machine learning, specifically in classification tasks, to rectify class imbalance. This occurs when one or more classes have notably lower sample sizes compared to others. When one class in a dataset has a disproportionately small number of instances compared to another, data balancing utilizing SMOTE (Synthetic Minority Over-sampling Technique) can compensate. The result of this imbalance can be machine learning models that favor the majority and underperform the minority.

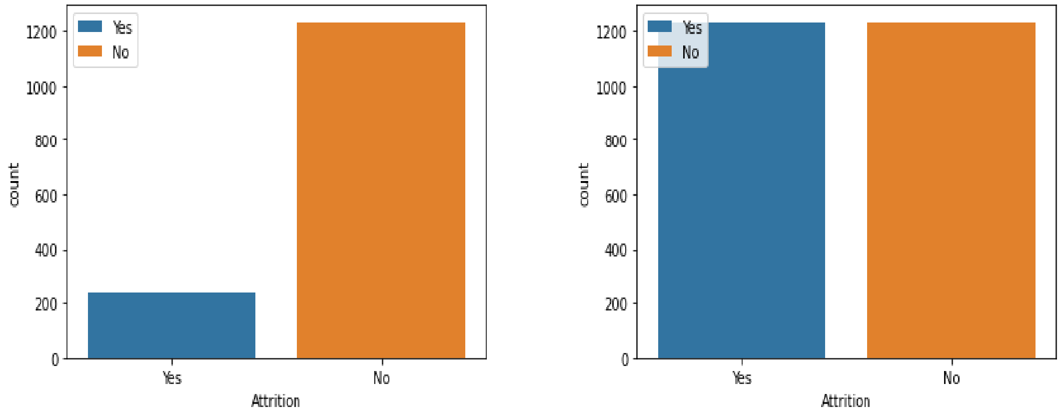

Figure 5: The dataset resampling using the SMOTE technique

Figure 5 displays two bar charts comparing the distribution of the "Attrition" variable before and after applying oversampling. In the first chart (left), the dataset is imbalanced, with significantly more "No" (non-attrition) cases than "Yes" (attrition) cases indicating a class imbalance issue. The second chart (right) shows the result after oversampling, where the number of "Yes" and "No" cases has been balanced to roughly equal levels. This technique helps improve model performance by ensuring that the classifier does not become biased toward the majority class.

3.7. Data Splitting

A dataset splitting was carried out, with the dataset being divided in half at the 85:15 ratio. A quarter of the dataset was used for testing purposes, while the other half was used for training the suggested machine learning model.

3.8. Proposed Deep Neural Network (DNN) Model

A DNN is a well-known deep learning method in the research community. In a DNN, the input, hidden, and output layers are all fully connected and form the network architecture. No neurones in one layer are connected to any neurones in another layer, even though all neurones in that layer are linked [24]. Each network layer's output undergoes activation function application to boost the network learning's impact. That being said, DNN is also somewhat similar to a large perceptron constructed from numerous smaller ones. This in action when the calculation for the forward propagation of the ith layer shown in Equation (2):

$$x_{i+1} = \sigma(\sum w_i x_i + b) \tag{2}$$

x stands for the input value, q for the weight coefficient matrices, and b for the bias vector. As an activation function, ReLU is commonly employed in multi-class networks. the equation (3) for this is:

$$\sigma(x) = max(0, x) \tag{3}$$

The loss function optimizes the network's backpropagation by evaluating the output loss of training samples and defining the network's structure. Equation (4) for cross-entropy, a loss function commonly used in classification applications is:

$$C = -\frac{1}{N}\sum_x \sum_{i=1}^{M}(y_i log p_i) \tag{4}$$

The amount of input data sets, the number of categories, the likelihood of forecasting into category i, and the chance that the classification i matches to the actual category are all variables in this equation. It utilized the following parameters to configure the Deep Neural Network (DNN): Adam optimizer (learning rate 0.001), ReLU activation, batch size 32, and 4,000 epochs. For binary classification, used cross-entropy loss; to avoid overfitting, used dropout = 0.2. Optimal training was guaranteed by early stopping based on validation loss and enhanced stability during initialization.

3.9. Evaluation Metrics

Various performance metrics were utilized to evaluate the effectiveness of the proposed methodology. The counts of True Negatives (TN), False Negatives (FN), True Positives (TP), and False Positives (FP) were arrived at by comparing the actual values with the anticipated outputs of the trained models. Here we'll go over how these numbers are used to calculate the four main evaluation metrics: recall, accuracy, precision, and F1-score:

Accuracy: The percentage of instances in the dataset (input samples) that the trained model accurately predicted as a unit of total occurrences. It is given as Equation (5).

$$Accuracy = \frac{\text{TP+TN}}{\text{TP+Fp+TN+FN}} \tag{5}$$

Precision: The ratio of the number of correctly predicted positive occurrences to the total number of positive occurrences predicted by the model is known as precision. Precision indicates. How good the classifier is in predicting the positive classes is expressed as Equation (6).

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{6}$$

Recall: This metric, the ratio of events that were accurately predicted as positive to all instances that should have proved positive. In mathematical form it is given as Equation (7).

$$Recall = \frac{\mathrm{TP}}{TP+FN} \tag{7}$$

F1 score: Together, they form the harmonic mean of memory and precision, which aids in maintaining a healthy equilibrium between the two. Its range is [0, 1]. Mathematically, it is given as Equation (8).

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

4. RESULTS AND DISCUSSION

This section gives a rundown of the experimental setup and shows the results of the suggested model's training and testing. A computer system with an AMD EPYC 7B12 type of central processor unit, 13 GB of random-access memory, a 2249.998 MHz CPU, and a 512KB cache size was used to run the suggested model. Important performance criteria utilized to train and evaluate the recommended model are accuracy, precision, recall, and F1-score (Table 2). Using the IBM HR dataset, the offered Deep Neural Network (DNN) model of employee attrition was evaluated, and all of the evaluation measures were found to be stable and effective. The overall performance of the model in identifying situations of employee attrition was high, with a degree of accuracy of 94.52%. With a precision of 94.58 %, it was also able to reliably identify the actual attrition instances with few false positives. Since the model's recall count was 94.52%, it correctly identified the majority of attrition cases. Additionally, the model's F1-score of 94.52% demonstrated that its precision and recall were functioning at parity. Such results demonstrate the efficacy and reliability of the DNN model for HR forecasting tasks.

Table 2: Experiment Results of dnn for Employee Attrition prediction on IBM HR dataset

Performance Matrix	**Deep Neural Network (DNN)**
Accuracy	94.52
Precision	94.58
Recall	94.52
F1-score	94.52

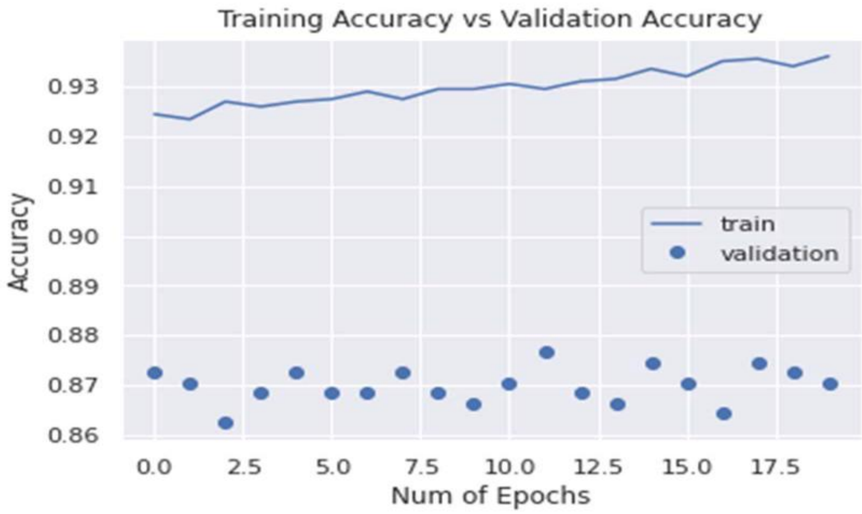

Figure 6: Accuracy curves for the DNN Model

In Figure 6, the accuracy rate of the proposed Deep Neural Network (DNN) is compared between the training and the validation processes, over 20 epochs. The training accuracy can be visualized through the line plot and it is constantly improving, with the initial point being approximately 92.3 percent and the final point being more than 93.5 percent, which is noteworthy because it means that the model is not stopping the training process and is not over training. Nevertheless, the validation accuracy is seen as a range of spontaneous blue dots at approximately 86-87.5%.

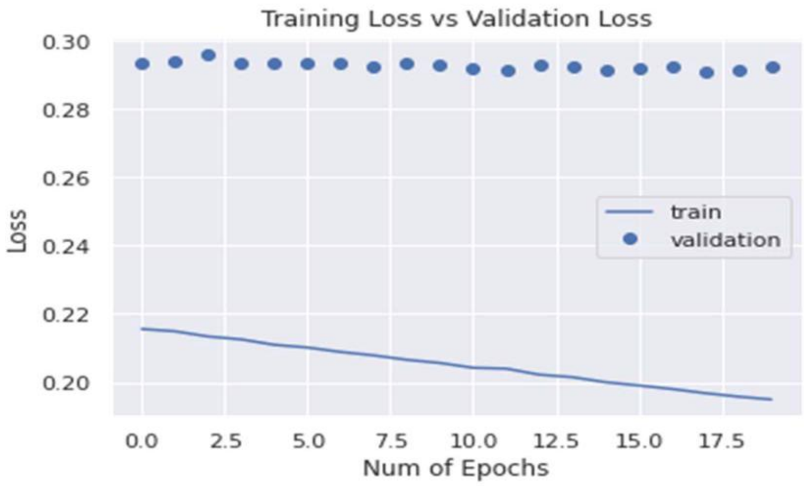

Figure 7: Loss curves for the DNN Model

The training and validation loss of the DNN during 20 epochs is shown in Figure 7. This indicates that the model is learning effectively from the training data, since the continuous line representing the training loss clearly decreases and continues to do so consistently at approximately 0.22, reaching a point where it balances between and below 0.19. Conversely, there is little change and no improvement in the validation loss, which is depicted by blue dots, across the epochs. Overfitting occurs when there is a continuous disparity between training and validation loss; this happens when a model fits the training data very well but fails to produce any relationship with the validation data.

4.1. Comparative Analysis

The effectiveness of the suggested DNN model was confirmed by carrying out accuracy comparison between the proposed model and other existing models, as presented in Table 3. The table in this assessment show the comparison of the accuracy of different predictive models used in making predictions using the employee attrition based on a resource-constrained HR system using the IBM HR dataset on Kaggle. The least accurate model was the LR which scored 74% showing little ability of predictive power. The best results were registered when using the RF model which did much better at 86.39% accuracy, followed by the best of all outcomes when using the SVM model where the accuracy reached 92%. The DNN with its highest model accuracy of 94.52% showed the best performance among all others due to having the capability of catching more intricate patterns and relationship in the data hence forecasting the most effective model in predicting attrition in study.

Table 3: Accuracy Comparison of different Predictive models of Employee Attrition prediction

Models	Accuracy
LR[25]	74
RF[26]	86.39
SVM[27]	92
DNN	94.52

The proposed Deep Neural Network (DNN) model shows that it represents a superior option since it predicts employee attrition with an accuracy of 94.52%. It automatically learns complex patterns in the data using its deep learning architecture and therefore is more effective compared to the traditional models. The accuracy of this is considered to be high which gives realistic forecasts on which the HR department can anticipate and design retention policies with more confidence.

5. CONCLUSION AND FUTURE STUDY

Employee turnover is a major issue to organizations that want to maintain the best talent in the new competitive global world. This paper proposes a DNN model which predicts the employee attrition based on the IBM HR Analytics dataset. The model proved to perform better as all the parameters accuracy, precision, recall, F1-score were averagely around 94.52%, even better than traditional ML models which include LR, RF, and SVM. The advantage of the DNN is the fact that it can learn complex nonlinear patterns, which is why it can be excellent at predicting possible resignation. To further facilitate the implementation of the model to HR purposes, explainable AI methods, such as SHAP or LIME, might be used to highlight the most influential factors that contribute to the rated factor of attrition to gain practical advice on reducing it. A real-time decision support system with the model deployed would enable the pro-active retention planning and improve the organizational stability. Besides, it would challenge the scalability and generalizability of the model to examine it with more mixed and bigger datasets and in other industries. Future work will be on how to combine explainable AI to work in the sphere of interpretability, cross-validating the model on a variety of industries and consider a hybrid or ensemble approach. By introducing time-series methods like LSTM or GRU, it will be possible to conduct predictions in a dynamic way to rely on employee behavioral patterns, which will enhance the ability of the model to make the predictions and be adaptable in real world.

REFERENCES

[1] T. Obaid *et al.*, “Factors Contributing to an Effective E- Government Adoption in Palestine,” in *Lecture Notes on Data Engineering and Communications Technologies*, 2022. doi: 10.1007/978-3-030-98741-1_55.

[2] B. S. Abunasser, M. R. J. Al-Hiealy, A. M. Barhoom, A. R. Almasri, and S. S. Abu-Naser, “Prediction of Instructor Performance using Machine and Deep Learning Techniques,” *Int. J. Adv. Comput. Sci. Appl.*, 2022, doi: 10.14569/IJACSA.2022.0130711.

[3] N. Mansor, N. S. Sani, and M. Aliff, “Machine Learning for Predicting Employee Attrition,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 435–445, 2021, doi: 10.14569/IJACSA.2021.0121149.

[4] Adarsh Reddy Bilipelli, “Application of AI and Data Analysis for Classification of Student Success in Large-Scale Educational Dataset,” *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 4, no. 6, pp. 428–441, Nov. 2024, doi: 10.48175/IJARSCT-22564.

[5] A. Elsafty and M. Oraby, “The Impact of Training on Employee Retention,” *Int. J. Bus. Manag.*, 2022, doi: 10.5539/ijbm.v17n5p58.

[6] A. Raza, K. Munir, M. Almutairi, F. Younas, and M. M. S. Fareed, “Predicting Employee Attrition Using Machine Learning Approaches,” *Appl. Sci.*, vol. 12, no. 13, 2022, doi: 10.3390/app12136424.

[7] R. Jain and A. Nayyar, “Predicting employee attrition using xgboost machine learning approach,” in *Proceedings of the 2018 International Conference on System Modeling and Advancement in Research Trends, SMART 2018*, 2018. doi: 10.1109/SYSMART.2018.8746940.

[8] H. Kali, “Diversity, Equity, and Inclusion Analytics in HR : How Workday Enables Data-Driven Decision-Making,” vol. 3, no. 2, pp. 162–170, 2023, doi: 10.56472/25832646/JETA-V3I6P113.

[9] M. Al Akasheh, E. F. Malik, O. Hujran, and N. Zaki, “A decade of research on machine learning techniques for predicting employee turnover: A systematic literature review,” *Expert Syst. Appl.*, 2024, doi: 10.1016/j.eswa.2023.121794.

[10] N. Prajapati, “The Role of Machine Learning in Big Data Analytics: Tools, Techniques, and Applications,” *ESP J. Eng. Technol. Adv.*, vol. 5, no. 2, pp. 16–22, 2025, doi: 10.56472/25832646/JETA-V5I2P103.

[11] S. N. Khera and Divya, “Predictive Modelling of Employee Turnover in Indian IT Industry Using Machine Learning Techniques,” *Vision*, 2019, doi: 10.1177/0972262918821221.

[12] A. R. Bilipelli, “End-to-End Predictive Analytics Pipeline of Sales Forecasting in Python for Business Decision Support Systems,” *Int. J. Curr. Eng. Technol.*, vol. 12, no. 6, pp. 819–827, 2022.

[13] W. Ding, M. Abdel-Basset, H. Hawash, and A. M. Ali, “Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey,” *Inf. Sci. (Ny).*, 2022, doi: 10.1016/j.ins.2022.10.013.

[14] F. Vaccari, N. Giovanelli, and S. Lazzer, “High-intensity decreasing interval training (HIDIT) increases time above 90% V̇ O2peak,” *Eur. J. Appl. Physiol.*, 2020, doi: 10.1007/s00421-020-04463-w.

[15] H. Kali, “The Future of HR Cybersecurity: AI-Enabled Anomaly Detection in Workday Security,” *Int. J. Recent Technol. Sci. Manag.*, vol. 8, no. 6, pp. 80–88, 2023.

[16] B. Tanmayi, H. Rishab, N. N. Yadav, A. A. S, and M. Venugopalan, “Comprehensive Analysis of Employee Attrition Using Machine Learning, Deep Learning and Ensemble Methods,” in *2025 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, 2025, pp. 1–6. doi: 10.1109/ICKECS65700.2025.11034779.

[17] V. T. R. P. K. M, V. Nagalakshmi, B. Sriram, S. A. Wasib, S. Sadiq, and M. D. Samhitha, “Improving Workforce Management with Machine Learning: A Novel Approach to Employee Classification,” in *2025 Third International Conference on Augmented Intelligence and Sustainable Systems (ICAISS)*, 2025, pp. 1411–1415. doi: 10.1109/ICAISS61471.2025.11042033.

[18] D. Muthugala, S. M. Arachchi, P. Pallewatta, A. Maithripala, and G. Seneviratne, “Predicting Employee Attrition & Employee Retention Period using Supervised Learning,” in *2024 6th International Conference on Advancements in Computing (ICAC)*, 2024, pp. 127–132. doi: 10.1109/ICAC64487.2024.10851009.

[19] N. S. Ozakca, A. Bulus, and A. Cetin, “Artificial Intelligence Based Employee Attrition Analysis and Prediction,” in *2024 6th International Conference on Computing and Informatics (ICCI)*, 2024, pp. 512–517. doi: 10.1109/ICCI61671.2024.10485157.

[20] A. Jin *et al.*, “Predicting Employee Attrition Using Machine Learning Approaches,” in *2024 9th International Symposium on Computer and Information Processing Technology (ISCIPT)*, 2024, pp. 419–426. doi: 10.1109/ISCIPT61983.2024.10672901.

[21] N. Silpa, V. V. R. Maheswara Rao, M. V. Subbarao, R. R. Kurada, S. S. Reddy, and P. J. Uppalapati, “An Enriched Employee Retention Analysis System with a Combination Strategy of Feature Selection and Machine Learning Techniques,” in *Proceedings of the 7th International Conference on Intelligent Computing and Control Systems, ICICCS 2023*, 2023. doi: 10.1109/ICICCS56967.2023.10142473.

[22] M. Maharana, R. Rani, A. Dev, and A. Sharma, “Automated Early Prediction of Employee Attrition in Industry Using Machine Learning Algorithms,” in *2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), ICRITO 2022*, 2022. doi: 10.1109/ICRITO56286.2022.9965017.

[23] R. Joseph, S. Udupa, S. Jangale, K. Kotkar, and P. Pawar, “Employee attrition using machine learning and depression

analysis," in *Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021*, 2021. doi: 10.1109/ICICCS51141.2021.9432259.

[24] R. Q. Majumder, "Machine Learning for Predictive Analytics: Trends and Future Directions," *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 4, pp. 3557–3564, 2025.

[25] A. Raza, K. Munir, M. Almutairi, F. Younas, M. Muhammad, and S. Fareed, "applied sciences Approaches," *Appl. Sci.*, 2022.

[26] K. Bhuva and K. Srivastava, "Comparative Study of the Machine Learning Techniques for Predicting the Employee Attrition," *Int. J. Res. Anal. Rev.*, vol. 5, no. 3, pp. 568–577, 2018.

[27] M. Tokmak, "DFCEA: Deep Forest Classifier-Based Employee Attrition Prediction," *Acta Infologica*, vol. 0, no. 0, pp. 0–0, 2025, doi: 10.26650/acin.1486319.

