Volume (1) No (12), 2025
Journal of Global Research in Multidisciplinary Studies (JGRMS)
Review Paper/Research Paper
Available online at https://saanvipublications.com/journals/index.php/jgrms/index

Employee Attrition Prediction through Machine Learning: Advancing Human Resource Analytics and Workforce Management

Dr. Nilesh Jain
Associate Professor
Department of Computer Sciences and Applications
Mandsaur University
Mandsaur
nileshjainmca@gmail.com

***Abstract*—The term "employee attrition" describes a certain reduction in a company's workforce due to a wide variety of factors. Employee attrition may also be caused by various factors in companies. Nevertheless, human resource managers must be able to identify signs of employee attraction at the initial stages. Organizational losses may be caused by employee attrition due to numerous reasons that include disruption of work and activities required to be done, re-employment and re-training cost, and loss of information. This study examines the accuracy of machine learning and deep learning models in forecasting employee turnover using the IBM HR Employee Attrition dataset. A number of models, including CNNs, were trained and evaluated against other methods, including DT, SVM, and LR. After comparing CNN model to more conventional machine learning algorithms, the experimental findings show that CNN model outperformed them with respect to accuracy (ACC) (92), precision (PRE) (96.67), recall (REC) (87), and F1-score (F1) (91.58). These findings indicate the efficiency of CNN to identify multifaceted trends in staff data, and it is a prospective HR analytics tool. These results demonstrate that predictive models based on deep learning have the potential to deliver actionable information to an organization to lower attrition and reinforce strategic human resource management.**

Keywords—Human Resources, Employee Turnover, Job Retention, Machine Learning, Human Resource Management, Employee Attrition, Attrition Prediction.

I. Introduction

In the modern competitive and fast-changing business world, the organizations in all industries are working towards ensuring stability, productivity and sustainable growth. One of the key factors behind these objectives is the workforce, which is usually considered as the best asset of any organization. It is direct contribution of the employees to the accomplishment of strategic goals, and companies spend enormous amounts of time and resources on selection of talented individuals, building their strengths, and creation of a productive working environment [1][2]. Nevertheless, it has been a growing challenge to retain a stable workforce with a number of internal and external pressures affecting employee behavior [3]. The high rate of staff turnover is a major problem for companies. Using analytics to make educated decisions and gain comprehensive insights into their staff is becoming increasingly important for human resources managers [4][5]. The attrition issue is bad when the employee turnover rate exceeds the rate of new talent acquisition where the organization is left with vacancies, poor productivity, and huge operational costs [6]. The reason is that the rate of attrition is a key that understands the performance and health of the organizations and the workforce [7][8].

As a reaction to those issues, companies are turning to human resource analytics (HRA), which is a data-based framework that provides insights into performance, engagement, workforce trends of their workers [9][10]. HRA helps HR departments understand the potential of the employees, their future requirements and methods to retain and develop employees [11][12]. Predictive analytics HR also enables organizations to identify the potential impact of multiple HR decisions and pre-plan interventions [13]. ML has emerged as a powerful tool of forecasting PRE in workforce attrition analysis with the complexity and volume of workforce information continuing to grow [14]. The ML algorithms can identify the latent trend on various employee data to generate plausible outcomes of the likelihood of employee exit [15]. This enable the HR leaders to initiate certain retention strategies, succession planning and developing organizational resilience.

A. Motivation and Contribution

The study's central argument is that organizations struggle to deal with employee turnover since it hinders their productivity, growth as an organization, and the cost of hiring new staff. Conventional approaches of studying attrition are not always able to reflect the multifaceted interaction of various conditions which make an employee choose to leave. As the data on HR and ML, as well as DL, become accessible, the possibility to create predictive models that able to identify an at-risk employee with good ACC emerges. This study become useful to organizations through the application of these methods, especially Convolutional Neural Networks (CNN), to offer relevant information that can be utilized to proactively manage attrition, improve employee retention efforts, and better manage the entire workforce. The study has some important contributions to the field of Employee Attrition:

- Used heatmaps, histograms, and pie charts to reveal feature distributions, correlations, and class imbalance, enhancing interpretability and transparency of the dataset.
- Achieving data parity and normalization using the SMOTE method to boost model performance.
- Introduced the application of a CNN to structured HR data undergo some modifications to its traditional application in image processing.
- Performed overall evaluation of model performance by several measures such as ACC, PRE, REC, ROC, and F1.

This study is justified by the fact that there is a growing need to have correct and sound predictive models in the management of human resources, where employee turnover directly affects the stability and performance of organizations. Although conventional ML models including DT, SVM, and LR have reported moderate success, they in many cases tend to overlook the non-linear associations between various employee characteristics. This study is novel in that the CNN is applied to structured HR data, which is not a common practice since CNNs are used to analyze image and spatial data. Modifying CNNs to predict employee attrition, this paper shows that they have the ability to learn complex feature representations automatically and enhance predictive ACC, and do so with higher results than the traditional models. This innovative procedure give organizations a better weapon of active attrition control and choice.

B. *Structure of Paper*

The following sections make up the paper: Section II is a literature review with an analytical gap. The suggested strategy that is used is outlined in Section III. Section IV contains the comparative and experiment outcomes. Lastly, Section V offers the conclusion, difficulties, and future research.

II. LITERATURE REVIEW

The development of this work has been guided and strengthened by a review and critical analysis of several important research works on employee attrition prediction.

Jeny et al. (2025) predicting the likelihood of employee turnover and primarily identifying causes of turnover are vital organizational objectives to optimize HR techniques. Interests, AI, ML, and DL have been actively applied to anticipate instability likelihood using automated techniques. This paper aims to utilize ML and DL models and compare them to attain the highest possible ACC. they intend to achieve an ACC of 94%, which is higher than the previous best of 92%, using the ANN and CNN algorithms [16].

Suddapally et al. (2025) proposes a new Multichannel Fuzzy Poplar Squeeze Excitation Network for employee attrition behavior prediction that can help with decision-making in the Employee Skills Inventory for Human Resource Management. The optimized network predicts the attrition behaviors, with fuzzy rules giving interpretable decisions. The model obtained ACC of 98.5%, PRE 99%, REC 98.5%, specificity 98.5%, f1-score 98.5%, MSE 0.15, MAPE 0.1, and R2 0.60 for attrition prediction. The suggested system enables HR managers to forecast possible attrition risks, enhance employee retention, and focus workforce management strategies on business goals [17].

Subha et al. (2024) suggests a new cascaded approach to predicting employee turnover that uses deep gradient optimization techniques to improve MLP and SVM models. The proposed methodology was tested against other models, including RF and LR. The cascaded MLP+SVM model had training and test accuracies of 94.5 and 92.1 percent respectively, as compared to standalone MLP and SVM, which had 89.7 and 86.5 percent respectively. Overfitting was also minimal in the model, as the difference between the training and test ACC was only 2.4% in the model. The model was proven to help with the classification of employees who were likely to leave because key performance metrics such as an F1 of 0.83 and an AUC score of 0.91 confirmed the ACC of the results of the classification [18].

Panda et al. (2024) the seven ML models are tested, with Extreme Gradient Boosting becoming the leader, scoring 84.4% on the test score and 84.5% on the F1 score. To extend the interpretability of the model further, they use feature importance analysis and this gives important understanding of factors affecting job change. By actively working to recognize the people at risk and take specific action using predictive information, the organizations able to create a more engaging and fulfilling working environment, which increase the retention rates, employee satisfaction and the overall reduction in employee turnover [19].

Mascarenhas, Savant and Aswale (2023) presents a research study that exploits ML methods to predict employee performance based on available data in organizations. Data collection, conditioning, and visualization are all part of the technique. Python classification methods including RF, SVM, KNN, DT, and NB are used for data classification. Using the ACC score and confusion matrix as metrics, Discover that the DT XGBoost algorithm achieved the best ACC of 82.7. However, when it came to actual positive results, the NB and SVM models performed better. This study offers useful understanding to the organizations that may want to use data to comprehend and solve the problem of attrition [20].

Arora, Kr Jha and Upadhyay (2023) entails training a model using an IBM analytics dataset on employee turnover, which contains 1,500 samples and 35 characteristics. Traditional metrics are employed to evaluate the model's efficacy following training. The supplied dataset yielded the most accurate results when the Gaussian Naïve Bayes classifier was applied. Its ability to correctly detect positive observations was demonstrated by its best REC (0.54) and its low false negative rate (4.5%) [21].

Maharana *et al.* (2022) study employs three ML models to forecast staff turnover based on the 35-feature IBM Watson dataset. This research aims to determine what causes voluntary employee turnover, what prevented certain employees from leaving, and how to utilize this information to forecast the likelihood of attrition. For this dataset, the LR ML method achieved the highest ACC(87%) and REC rate (0.36), two performance metrics used to summarize the results [22].

A literature study revealed that ML, DL, and hybrid models have made significant progress in predicting employee attrition. However, these models still suffer from a number of limitations that limit their usefulness and scope of application. As per the summary of investigations summarized in Table I, the majority of the current literature depends on the IBM HR dataset to a large extent, raising questions about the validity of the model in different organizational settings. Some of them are very accurate, like the Multichannel Fuzzy Poplar Squeeze Excitation Network and cascaded MLP-SVM, but many of them cannot be interpreted, which is essential to HR decision-making. There are also some other problems like class imbalance, low recalls, low emphasis on real-time prediction, inadequate inclusion of behavioral, organizational or external factors which still do not support much in predictive depth. Additionally, the majority of researches focus on PRE but not on applicability in HR practice, integration of employee sentiments, and practical retention techniques. These gaps imply that the more explainable, generalized, and operationally deployable predictive frameworks are needed.

TABLE I. OVERVIEW OF RECENT STUDIES ON PREDICTIVE MODELING OF EMPLOYEE ATTRITION PREDICTION

Author	Proposed Work	Dataset	Key Findings	Challenges/Recommendations
Jeny et al., (2025)	Utilized ML and DL models (ANN, CNN) to predict employee turnover	IBM HR Dataset	Achieved 94% accuracy, higher than previous 92%; emphasized identifying causes of attrition	Focused on optimizing accuracy; further research needed to integrate interpretability of model outputs
Suddapally et al., (2025)	Multichannel Fuzzy Poplar Squeeze Excitation Network for attrition prediction	Employee Skills Inventory dataset	Accuracy 98.5%, Precision 99%, Recall 98.5%; fuzzy rules provided interpretable decisions	Complexity of model; recommendation to simplify for practical HR adoption
Subha et al., (2024)	Cascaded MLP + SVM enhanced by deep gradient optimization	IBM HR Dataset (preprocessed with SMOTE)	Training accuracy 94.5%, Test accuracy 92.1%; F1 score 0.83, AUC 0.91; minimal overfitting	Requires careful feature selection; addressing overtime and job satisfaction in retention strategies
Panda et al., (2024)	ML pipeline for job change prediction using seven models	"HR Analytics: Job Change of Data Scientists" Kaggle dataset	XGBoost test score 84.4%, F1 score 84.5%; feature importance revealed key job change factors	Class imbalance challenges; recommendation for Proactive Intervention Strategies
Mascarenhas, Savant and Aswale (2023)	ML-based employee performance and attrition analysis using RF, SVM, KNN, DT, Naive Bayes	Organization HR dataset	Decision Tree XGBoost achieved 82.7% accuracy; Naive Bayes and SVM better in true positives	Need to reduce false negatives; focus on key attrition causes for actionable insights
Arora, Kr Jha and Upadhyay (2023)	Data-driven model to forecast employee leaving using Gaussian Naïve Bayes	IBM HR Analytics dataset (35 features, ~1500 samples)	Best recall 0.54; low false negatives (4.5%); accurately identified employees likely to leave	Low recall indicates room for improvement in detecting all potential leavers
Maharana et al., (2022)	ML models for attrition prediction on IBM Watson dataset (3 models)	IBM Watson HR Dataset (35 features)	Logistic Regression achieved 87% accuracy and best recall 0.36	Limited recall; recommendation to enhance predictive power and retention strategies

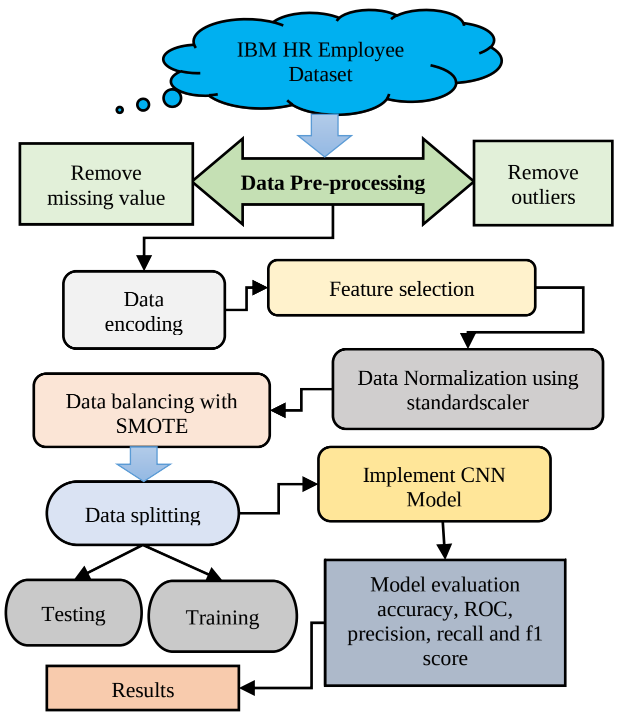

Fig. 1. Proposed flowchart for Employee Attrition Prediction

III. RESEARCH METHODOLOGY

The methodology of the study included extracting insights on the reasons and sources of employee turnover using IBM HR's Employee Attrition data. Preprocessing of the data involved the process of dealing with missing values, outliers as well as the categorical variables, feature selection, normalization and balancing classes using SMOTE. A portion of it (85%) was used for training, and a smaller portion (15%) was used for testing. Using max pooling, flattening, and thick layers, a suggested CNN model was developed to categorize attrition outcomes. Multiple studies were conducted to assess the model's performance, including ACC, PRE, REC, F1, and ROC. These were chosen due to their high prediction capabilities and their ability to properly represent both the attrition and non-attrition classes. Show the entire procedure in Figure 1.

A detailed explanation of each step in the proposed flowchart for employee attrition prediction is provided below.

A. Data collection

The IBM HR Employee Attrition dataset was used to build a generalized ML model that can forecast the attrition of valuable employees. The creation of the data was carried out by data scientists at IBM. The dataset has 35 features. Through analysis of the dataset, and were able to identify the causes of employee turnover. The collection comprises 1470 employee records. Data visualizations such as bar plots and heatmaps were used to examine Employee Attrition Prediction, feature correlations, etc., are given below:

The Pearson correlation coefficients of several employee-related characteristics on the goal variable attrition are displayed in the heatmap in Figure 2. There is a weak relationship between attrition and age, total working years, years in present position, and years with current boss, according to the data. Longer tenure and experience seem to be linked to a decreased chance of quitting, according to this. In contrast, attributes with modest positive connections, such as percent salary hike and number of companies worked, are not very influential. And may use this visualization to find out which factors are the most important to include as features in model for predicting employee attrition.

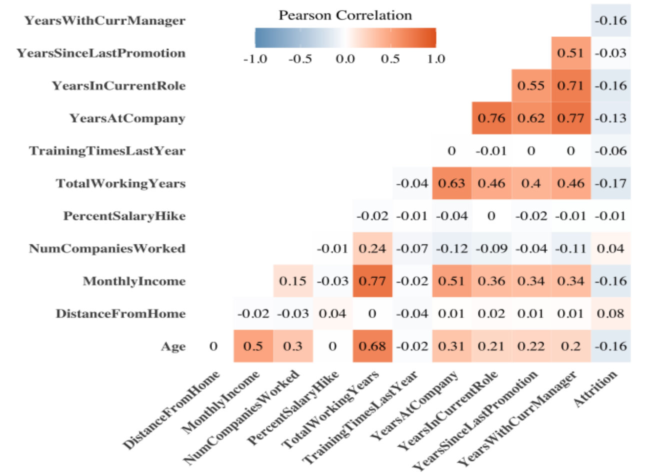

Fig. 2. Correlation analysis for features

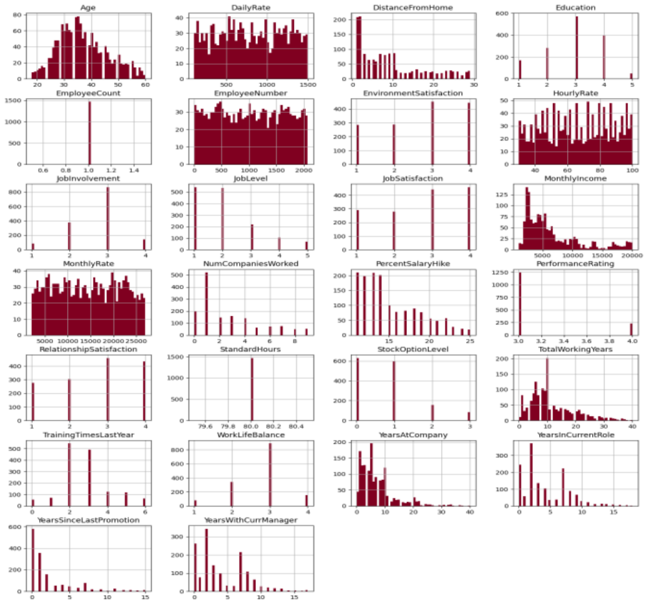

Fig. 3. Histogram for attribute frequency

The grid of histograms in Figure 3 illustrates the frequency of 25 attributes involving employees in the dataset. These visualizations have the potential to provide light on the distribution and concentration of values for a variety of important characteristics, including age, monthly income, job satisfaction, total working years, and distance from home. Using these distributions, analysts can determine skewed characteristics, detect possible outliers and determine the quality of data before modeling. This summary is useful in choosing features and preprocessing predictive analytics.

B. Data Pre-Processing

Using the IBM HR Employee Dataset, merging and cleaning the data, and extracting related characteristics were all steps in the data preparation process. Prior to data transformation and normalization, preprocessing operations were implemented to address outliers and missing values. Here are the main steps of pre-processing:

- **Remove missing value:** Handling missing data can be done in several ways based on the type of data and level of missingness, but it is better to remove missing values.
- **Remove Outliers:** The term "outlier" refers to data points in a dataset that are extremely out of the ordinary. Finding these outliers and handling them, often by reducing, replacing, or modifying the data to lessen their influence, is known as "outlier removal."

C. Data Encoding

The dataset included numerically coded category variables that may be utilized by ML methods. I labelled each category using integer values using the Label Encoder technique. The values of the Business Travel column were classified as 0, 1, and 2 for the cases where the travel frequency was low, medium, and high, respectively. Gender was categorized as male and female, for the other cases.

D. Feature Selection

The aim of feature selection is to enhance the accuracy, predictability, and interpretability of a ML model by selecting the most relevant characteristics from a dataset. The feature selection process removes non-informative or unimportant features and decreases noise, simplifies the model, avoids overfitting, and reduces the computational expenses, in turn, makes feature predictions more efficient and reliable. The process of selecting a subset of relevant features (variables or attributes) in a dataset out of a larger set of characteristics of the dataset is called feature selection in data preprocessing.

E. Data Normalization using standardscaler()

The data is changed using the StandardScaler() method so that the mean is 0 and the standard deviation is 1. By subtracting the average value from each observation and dividing by the standard deviation, as shown in Equation (1), and may achieve this translation:

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

Where x represents the starting point for every descriptor, z the converted feature value, μ the mean of the collection, and σ its standard deviation.

F. Data Balancing Using SMOTE

Data balancing approaches correct class imbalance in datasets, providing fair representation and boosting model performance for minority classes. The dataset was resampled using the SMOTE method. Dataset balancing reduced the model's complexity. Improved ACC is the result of training the model with an equal number of target distributions.

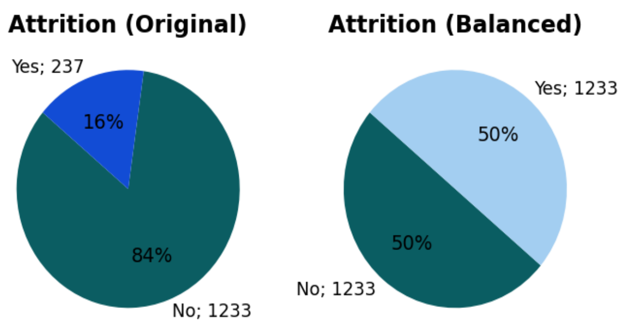

Fig. 4. Pie Chart for Original and Balanced Dataset Distribution

The comparison between the initial and the balanced employee attrition data is presented in Figure 4. The Original, shows a strong imbalance in the classes, with 16 percent of employees with the Yes to attrition and 84 per cent with No, showing a skewed distribution that may be discriminatory to predictive models. The Balanced, on the other hand, presents the same share of both classes, Yes and No, but it has been done in data balancing methods. This visual analogy highlights the necessity to consider the imbalance of classes in ML processes in order to guarantee just training of models and make sure that they perform well in minority and majority classes.

G. Data Splitting

The generalization of model was enhanced by the dataset splitting. At 85:15, the split was executed. The model was evaluated using 15% of the data, after being trained on 85%.

H. Proposed Convolutional Neural Network (CNN) Model

The processing and recognition of pictures are common applications of CNNs, a type of ML method that uses an artificial neural network to detect patterns in images [23]. A CNN consists of several layers, including convolutional, pooling, and fully connected ones. The pooling layer removes the input parameters after the convolutional layer has converted the picture to numerical values. When it comes to detection and processing of images, the DL technique known as CNN is ideal. Multiple layers make it up, including completely connected, pooling, and twisted layers. Figure 5 provides a high-level representation of the proposed CNN model for predicting employee turnover.

Model: "sequential_1"

Layer (type)	Output Shape	Param #
conv2d_3 (Conv2D)	(None, 126, 126, 32)	896
max_pooling2d_2 (MaxPooling2D)	(None, 63, 63, 32)	0
conv2d_4 (Conv2D)	(None, 61, 61, 64)	18,496
max_pooling2d_3 (MaxPooling2D)	(None, 30, 30, 64)	0
conv2d_5 (Conv2D)	(None, 28, 28, 128)	73,856
max_pooling2d_4 (MaxPooling2D)	(None, 14, 14, 128)	0
flatten (Flatten)	(None, 25088)	0
dense_2 (Dense)	(None, 128)	3,211,392
dense_3 (Dense)	(None, 1)	129

Total params: 3,304,769 (12.61 MB)
Trainable params: 3,304,769 (12.61 MB)
Non-trainable params: 0 (0.00 B)

Fig. 5. Summary of Proposed CNN Model

The model begins with three convolutional layers (Conv2D) with varying depths—32, 64, and 128 filters and uses max pooling to down-sample the spatial dimensions in between. Two dense layers, one with 128 units for feature abstraction and the other with 3 units for final output, are added to the network after the feature maps are flattened. These layers are used for classification. This architecture is the balance between depth and efficiency, with more than 3.3 million trainable parameters, which is adequate to much less complex visual tasks.

I. Evaluation Metrics

A number of performance metrics were used to assess the effectiveness of the proposed design. Compared the trained models' actual values to their predicted outcomes. The TN, FN, TP, and FP were calculated using this comparison. Named here are the components of the following matrix: REC, ACC, PRE, and F1:

Accuracy: A metric for evaluating the trained model's predictive performance relative to the whole dataset (input samples). It is stated as Equation (2)-

$$Accuracy = \frac{TP+TN}{TP+Fp+TN+FN} \quad (2)$$

Precision: The term "precision" refers to the ratio of the number of positive cases that the model accurately predicts to the total number of positive examples. Accuracy may be expressed as Equation (3):

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Recall: The ratio of positively expected occurrences to all instances that ought to have been positively impacted is this metric. It is expressed mathematically as Equation (4)-

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

F1 score: A combination of the harmonic mean of REC and PRE, it helps to bring REC and PRE into harmony with each other. [0, 1] is its range. It is expressed mathematically as (5).-

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Receiver Operating Characteristic Curve (ROC): The ROC plot compares, across a range of decision cut-off points, the proportion of instances appropriately and incorrectly categorized as positive. In contrast to TPR, which is also called sensitivity or recall, FPR is equivalent to 1-specificity.

IV. Results and Discussion

Utilizing a Jupyter Notebook (version: 7.3.2) environment and Python (version: 3.12.7), the experimental research made use of the suggested model. An Intel 13th-generation Core i9 13900K, 128 GB of RAM, and an NVIDIA GeForce RTX 4080 SUPER GPU were the components of the Windows 11 PC used in this investigation. The proposed model was assessed using the primary performance measures of ACC, PRE, REC, and F1 and trained on the IBM HR Employee Dataset. The results are summarized in Table II. The model was able to reach a high ACC of 92 which means that there was a high degree of ACC in the predictions overall. The CNN has an excellent reliability when it comes to the identification of the right employees who are expected to leave the company with a PRE of 96.67, thereby reducing false positives. These findings reveal the efficiency of the CNN in dealing with complicated trends in HR data, which makes it a very appropriate model to make precise predictions of workforce attrition.

TABLE II. Experiment Results of Proposed Models for Employee Attrition Prediction on IBM HR Employee Dataset

Performance Matrix	Convolutional Neural Network (CNN) Model
Accuracy	92
Precision	96.67
Recall	87
F1-score	91.58

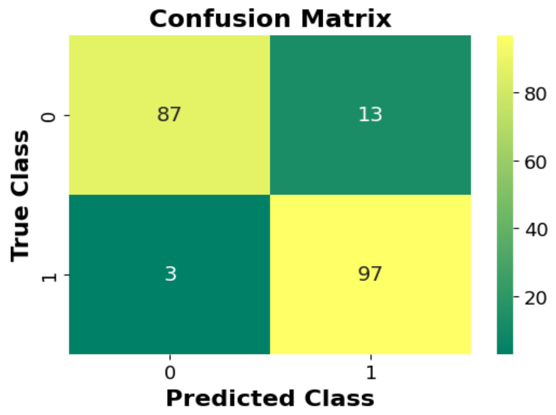

Fig. 6. Plot for Confusion matrix for CNN

The detail of the model performance in terms of classification is depicted in Figure 6. The model was highly accurate by classifying 97 cases correctly and only 3 falsely, indicating a high ACC on the negative category among the true class 0 cases. In the case of the true class 1, it made the right classification of 87 cases and wrongly classified 13 cases as class 0. Generally, the model has a high predictivity rate with many correct classifications in both classes, but little skew in identifying class 1, which has the potential to further enhance the REC of the positive class.

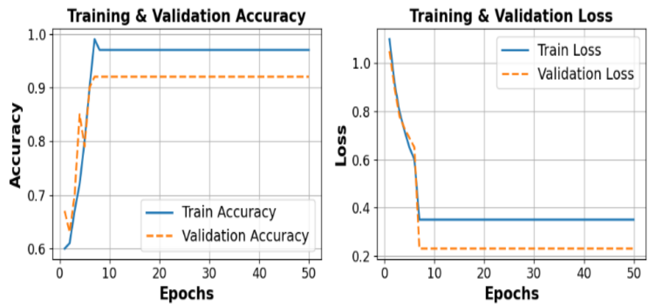

Fig. 7. Loss and Accuracy Curve of CNN Model

Figure 7 indicates that the CNN model achieves ACC values of 100% and 92% in the first several epochs, respectively, during training and validation. The two loss curves also drop off drastically and flatten out, which is an indication of good learning and minimum overfitting. The consistency of trends in training and validation testifies to the consistency of model performance, and effective generalization.

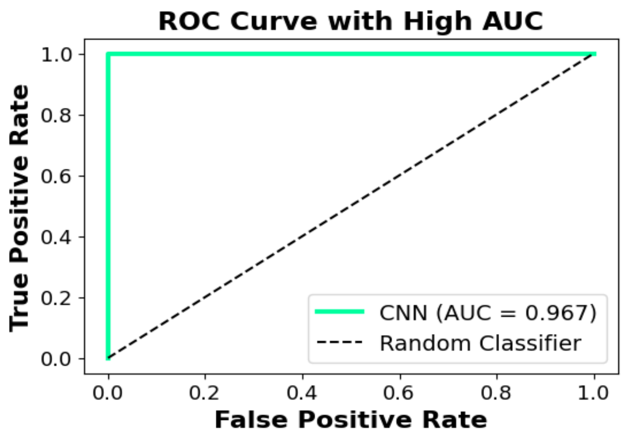

Fig. 8. ROC analysis of the CNN model

Figure 8 shows that it can differentiate between classifications that are positive and negative. A steepening of the curve towards the top left corner indicates a low false positive rate and a high genuine positive rate at different thresholds. This model has excellent discriminating power and a high AUC of 0.96, indicating that it is good at data classification. The high value of this AUC value is an assurance that the CNN is very good at identifying complicated patterns in the data and it is capable of giving good predictions when it comes to tasks related to fraud or attrition detection.

A. Comparative Analysis

The efficiency of the proposed CNN model was verified by a comparison examination of ACC with other existing models, as shown in Table III. The DT model was 79.4% accurate with all PRE, REC, and F1 values of 79.5 which can be described as moderately accurate but with less ability to generalize. SVM had an ACC of 87% with good consistency in ACC, REC and F1, LR had an ACC of 89%, but the F1 significantly decreased to 31.26% indicating that the system lacked balance in its ability to REC and PRE with PRE, probably because of the problem of unequal classes. The CNN however proved superior to all the models and showed its capability to identify more intricate patterns and make more accurate and balanced predictions of employee attrition. This comparison makes CNN the best model to improve human resource analytics and workforce management.

TABLE III. COMPARISON OF DIFFERENT PREDICTIVE MODELS ON IBM HR EMPLOYEE DATASET

Model	Accuracy	Precision	Recall	F1-Score
DT[24]	79.4	79.5	79.5	79.4
SVM[25]	87	87	87	87
LR[26]	89	-	-	31.26
CNN	**92**	**96.67**	**87**	**91.58**

The CNN model that is being advertised has its unique advantages and innovativeness in terms of forecasting employee turnover. As compared to the traditional ones such as DT, SVM, and LR, the CNN can be used to identify complex and non-linear trends among HR data, which are more precise, balanced in terms of ACC, REC, and F1. Fairness and generalization is further increased by the addition of powerful preprocessing measures (feature selection, normalization, and SMOTE balancing). Such a combination of a new CNN application and thorough data preparation creates a new and effective framework of workforce analytics and prediction of attrition.

V. CONCLUSION AND FUTURE STUDY

Analytics is described as a methodical progression and a set of statistical tools. To put it another way, it's the science of analysis. HR analytics is a methodical discovery and quantification of the people-related aspects that drive business results. It involves demonstrating the direct influence of people data on crucial business outcomes. HR analytics is positioned as a critical requirement for businesses and forecasts employee attrition. This study introduced a new CNN model that can predict employee turnover using the IBM HR Employee Attrition dataset. The dataset was narrowed down to fairness and reliability, through extensive preprocessing, such as missing values, outliers, feature-selection, and normalization, and balancing class representation with SMOTE. It was experimentally proved that the CNN model has a better performance (ACC: 92, PRE:

96.67, REC: 87, F1: 91.58, AUC: 0.96) in comparison with the traditional ML model, including DT, SVM, and LR. The capability of the CNN to record the complex and non-linear relationships in HR data demonstrates its novelty and usefulness in workforce analytics. The article is also constrained by the use of a single dataset (which might limit the generalizability) and the possibility of a bias due to the use of synthetic samples generated by SMOTE. It could be confirmed in future work that the model is tested on a variety of, more realistic, HR datasets, that the hybrid DL methods are tested, and that explainable AI is used to enhance interpretability and practical application.

References

[1] A. Elsafty and M. Oraby, "The Impact of Training on Employee Retention," *Int. J. Bus. Manag.*, vol. 17, no. 5, p. 58, Apr. 2022, doi: 10.5539/ijbm.v17n5p58.

[2] 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.

[3] A. Parupalli, "Business-Oriented Employee Performance Assessment via Machine Learning in ERP Systems," *Tijer – Int. Res. J.*, vol. 11, no. 11, 2024.

[4] S. F. Sari and K. M. Lhaksmana, "Employee Attrition Prediction Using Feature Selection with Information Gain and Random Forest Classification," *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 410–419, Sep. 2022, doi: 10.47065/josyc.v3i4.2099.

[5] 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.

[6] S. Najafi-Zangeneh, N. Shams-Gharneh, A. Arjomandi-Nezhad, and S. H. Zolfani, "An improved machine learning-based employees attrition prediction framework with emphasis on feature selection," *Mathematics*, vol. 9, no. 11, 2021, doi: 10.3390/math9111226.

[7] S. R. Ponnuru, "Employee Attrition Prediction using Logistic Regression," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 8, no. 5, pp. 2871–2875, May 2020, doi: 10.22214/ijraset.2020.5481.

[8] R. Punnoose and P. Ajit, "Prediction of Employee Turnover in Organizations using Machine Learning Algorithms," *Int. J. Adv. Res. Artif. Intell.*, vol. 5, no. 9, 2016, doi: 10.14569/IJARAI.2016.050904.

[9] M. Nandal, V. Grover, D. Sahu, and M. Dogra, "Employee Attrition: Analysis of Data Driven Models," *EAI Endorsed Trans. Internet Things*, vol. 10, Jan. 2024, doi: 10.4108/eetiot.4762.

[10] S. N. Mishra, D. R. Lama, and Y. Pal, "Human Resource Predictive Analytics (HRPA) For HR Management In Organizations," *Int. J. Sci. Technol. Res.*, 2016.

[11] V. Kakulapati, K. K. Chaitanya, K. V. G. Chaitanya, and P. Akshay, "Predictive analytics of HR - A machine learning approach," *J. Stat. Manag. Syst.*, vol. 23, no. 6, pp. 959–969, Aug. 2020, doi: 10.1080/09720510.2020.1799497.

[12] S. P. Kalava, "Revolutionizing Customer Experience: How CRM Digital Transformation Shapes Business," *Eur. J. Adv. Eng. Technol.*, vol. 11, no. 3, pp. 163–166, 2024.

[13] 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.

[14] S. Thangavel, S. Srinivasan, S. B. V. Naga, and K. Narukulla, "Distributed Machine Learning for Big Data Analytics: Challenges, Architectures, and Optimizations," *Int. J. Artif. Intell. Data Sci. Mach. Learn.*, vol. 4, no. 3, pp. 18–30, Oct. 2023, doi: 10.63282/3050-9262.IJAIDSML-V4I3P103.

[15] 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, May 2025, doi: 10.38124/ijisrt/25apr1899.

[16] J. R. V Jeny, V. S. Chowdhary, A. N. Reddy, and P. N. Goud, "Detection and Prevention of Workflow Attrition," in *2025 5th International Conference on Pervasive Computing and Social Networking (ICPCSN)*, IEEE, May 2025, pp. 1567–1573. doi: 10.1109/ICPCSN65854.2025.11035430.

[17] L. Suddapally, R. kumar J, P. K. Parida, D. Barani, and G. S. Doguparthy, "Deep Learning-Based Employee Skills Inventory and Attrition Prediction for Human Resource Management," in *2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL)*, 2025, pp. 700–706. doi: 10.1109/ICSADL65848.2025.10933165.

[18] T. D. Subha, S. Vijayalakshmi, M. N. Pooja, R. K. V, J. Pavithira, and R. V, "Experimental Evaluation of a Novel Methodology to Predict Employee Attrition Using Elevated Deep Gradient Optimization Logic," in *2024 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, 2024, pp. 1–6. doi: 10.1109/ICSES63760.2024.10910261.

[19] A. R. Panda, D. Sanyal, A. Sur, and M. K. Mishra, "Attention to Retention: Exploring the Dynamics of Data Scientist Job Mobility," in *2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET)*, 2024, pp. 1–6. doi: 10.1109/ACROSET62108.2024.10743566.

[20] J. Mascarenhas, S. Savant, and S. Aswale, "Employee Appraisal Prediction Using Classification Models," in *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)*, 2023, pp. 2170–2175. doi: 10.1109/IC3I59117.2023.10398114.

[21] S. Arora, A. Kr Jha, and S. Upadhyay, "Predicting a Rise in Employee Attrition Rates Through the Utilization of People Analytics," in *Proceedings of the 2023 12th International Conference on System Modeling and Advancement in Research Trends, SMART 2023*, 2023. doi: 10.1109/SMART59791.2023.10428268.

[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] S. R. Sagili, S. B, P. RVS, and A. P, "Sentiment Classification for Depression Detection: Integrating Capsule Networks with CNNs on Review Data," in *2025 Emerging Technologies for Intelligent Systems (ETIS)*, IEEE, Feb. 2025, pp. 1–7. doi: 10.1109/ETIS64005.2025.10961685.

[24] M. Tokmak, "DFCEA: Deep Forest Classifier-Based Employee Attrition Prediction," *Acta Infologica*, vol. 9, no. 1, pp. 19–33, Feb. 2025, doi: 10.26650/acin.1486319.

[25] 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.

[26] F. Guerranti and G. M. Dimitri, "A Comparison of Machine Learning Approaches for Predicting Employee Attrition," *Appl. Sci.*, vol. 13, no. 1, p. 267, Dec. 2022, doi: 10.3390/app13010267.

