Journal of Information Systems Engineering and Management
2025, 10(2)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Employee Attrition Classification using Improved Deep Belief Networks

[1]A. Senthilvelan, [2]Dr. M. Sengaliappan,
[1]*Research Scholar, Nehru College of Management, Coimbatore*
[2]*Head, Department of M.C.A., Nehru College of Management, Coimbatore*

ARTICLE INFO

Received: 28 Oct 2024

Revised: 05 Jan 2025

Accepted: 22 Jan 2025

ABSTRACT

Companies are often worried about employee turnover since it affects productivity, morale, and expenses. It is common for traditional attrition prediction approaches to be inaccurate and to not make full use of the data that is available. Using Improved Deep Belief Networks (IDBN), we provide a state-of-the-art method for employee turnover classification in this research. The first of the two steps in the approach is feature selection, which is used to choose the best variables to use for prediction. Then, using this improved dataset, the IDBN model is trained to correctly identify stable and attrition-prone personnel. Integrating IDBN improves the model's capacity to detect intricate patterns and connections in the data, which is the main contribution of this study. To top it all off, the feature selection process guarantees that the model uses high-quality inputs, which boosts classification accuracy even further. The suggested strategy outperforms conventional approaches and achieves better accuracy in forecasting employee turnover, as shown experimentally using real-world datasets. If your company is looking to optimize worker stability and proactively control attrition, this study gives significant information.

Keywords: Classification, Deep learning, Employee attrition, Feature selection, Improved Deep Belief Networks.

I. Introduction

Employee attrition refers to the process by which a company loses its personnel for many reasons, including but not limited to: poor working conditions, low pay, dissatisfaction with one's own personal life, and negative company culture. There are two main types of employee turnover: voluntary and involuntary [1]. Involuntary attrition occurs when employees are fired by their employers due to factors like poor performance or business requirements [2]. In contrast, voluntary attrition occurs when high-performing employees willingly leave their employment despite employers' best attempts to retain them [3]. Voluntary attrition can take several forms, including early retirement and job offers from other companies [4]. Voluntary attrition and the loss of brilliant individuals affect even organizations that value their employees and invest in them via outstanding working conditions and extensive training [5]. The expense of recruiting, employing, and training new employees is another major concern when it comes to replacing departing employees [6].

A company will always incur substantial expenses if it hires new staff. The time and money spent on training a new employee and seeing them through to full productivity are examples of measurable revenues [7-8]. Every day, HR departments create a mountain of data from things like employee absences, social disputes, yearly reviews, salaries, benefits, new hires, firings, career reviews, and more [9–10]. Finding the right and appropriate replacements for the departing personnel is the major challenge, but [11]. Management can improve internal policies and initiatives in response to employee turnover rate predictions [12, 13]. Where there are a number of options available to bright workers who are considering leaving, such as a pay rise or better training, in order to keep them from leaving [14-15]. Businesses can anticipate staff turnover with the use of machine learning models. Analysts can construct and train a machine learning model to forecast which workers will leave the organization using the historical data stored in HR departments. During training, these models look for commonalities between current and former workers' characteristics [16–18].

The main contribution of the paper is

- Classification using improved Deep Belief Networks

This paper is organized as follows for the rest of it. Part 2 of the book covers a wide range of writers' approaches on predicting employee turnover. Finally, in Section 3, we see the suggested model. The findings of the study are reviewed in Section 4.

1.1 Motivation of the paper

The motivation behind this paper is to address the pressing issue of employee attrition, which significantly impacts organizational productivity, morale, and costs. Traditional methods for attrition prediction often fall short in accuracy and fail to harness the full potential of available data. Therefore, we propose an advanced approach using Improved Deep Belief Networks (IDBN) combined with feature selection to enhance the accuracy and effectiveness of employee attrition classification. This research aims to provide organizations with valuable insights and tools to proactively manage attrition, optimize workforce stability, and make data-driven decisions for sustainable business outcomes.

II. Background study

Alshiddy, M. S., & Aljaber, B. N. [1] A key worry that impacts a company's success was an increase in personnel turnover rates. The success of the hiring process has an effect on it since a lower turnover rate was associated with an improved hiring process. This article explored the impact of using layered ensemble learning on enhancing employee attrition prediction.

El-Rayes, N. et al. [3] Overall, the author were able to collect a dataset of employees' job changes using Glassdoor's web portal and resumes submitted anonymously. A preliminary analysis of this data set yielded various discoveries, the most important of which were the importance of pay, corporate culture, and the performance of upper management in determining whether or not an employee would decide to leave their current position. After that, the author dug deeper into this data by creating a job transition table for the industry, finding out which variables had the most shifts in distribution between workers who remained and those who departed, and building rating features from a principal component analysis (PCA) research.

Jain, D. [5] these authors research used a hybrid model of ensemble techniques based on the CRISP-DM business model to conduct predictive assessments on employee attrition for IBM, a major technology firm. The analysis focused on effective feature selection. Adaptive boosting had the highest assessment scores with an accuracy of up to 88.8 percent, while stacking and bagging were the most effective ensemble approaches for classification.

Kumari, S. [7] In order to address the study issue and achieve the research goals, it was necessary to fill a vacuum in the area of employee attrition prediction, as was evident from the literature evaluation. The process for creating a system to forecast staff turnover was detailed in the next chapter.

Özdemir, F. [11] In order to accurately anticipate employee attrition, the study set out to conduct a number of data transformations, including feature engineering, feature encoding, and scaling, and to choose from a variety of methods for tuning the hyperparameters on classification models. Organisations might use the research's practical implications to improve their staff management strategies and reduce the negative effects of employee turnover. The elements that hint to the likelihood of attrition were uncovered by doing exploratory data analysis on the relevant data. Employees' motivation observed to decrease when monetary considerations such monthly income, daily rate, and hourly rate were included.

Yadav, S. [17] an organization's goodwill, earnings, and time and money costs can all take a hit when employees leave. In addition to assisting with preventative measures, the predicted attrition model also aids in making more informed recruiting choices. This research uses a variety of categorization methods to forecast, using patterns in an employee's historical data, if that individual was likely to quit the company soon. Employees leave for reasons other than pay or other monetary considerations, such as promotions.

Table 1: Comparison table existing work

Author	Year	Methodology	Advantage	Limitation
Alshiddy & Aljaber	2023	Nested Ensemble Learning Techniques	Combines multiple models for improved accuracy and robustness	Complexity in model training and interpretation
El-Rayes et al.	2020	Tree-Based Models	Easy to interpret and handle both numerical and categorical data	Can be prone to overfitting and can require pruning
Jain et al.	2020	Machine Learning Approach	Provides insights into employee attrition factors with high prediction accuracy	Requires a large dataset for training and can suffer from bias
Mohbey	2020	Machine Learning Approaches	Real-time application capability with various ML algorithms	Requires significant computational resources and data preprocessing
Veerkhare et al.	2022	Modern ML Approach	Innovative use of modern machine learning techniques for better performance	Can face challenges with data privacy and model transparency

2.1 Problem definition

The problem addressed in this study is the challenge of accurately predicting and managing employee attrition within organizations. Employee attrition, which refers to the rate at which employees leave a company, has significant implications for organizational performance, including impacts on productivity, morale, and costs. Traditional methods of attrition prediction often fall short in terms of accuracy and fail to leverage the full potential of available data.

III. Materials and methods

In this section, we detail the proposed methodology for Employee Attrition Classification using Improved Deep Belief Networks (IDBN). The methodology comprises a two-phase process, beginning with feature selection to identify the most relevant variables for prediction.

3.1 Dataset collection

The dataset was collected from Kaggle website https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset

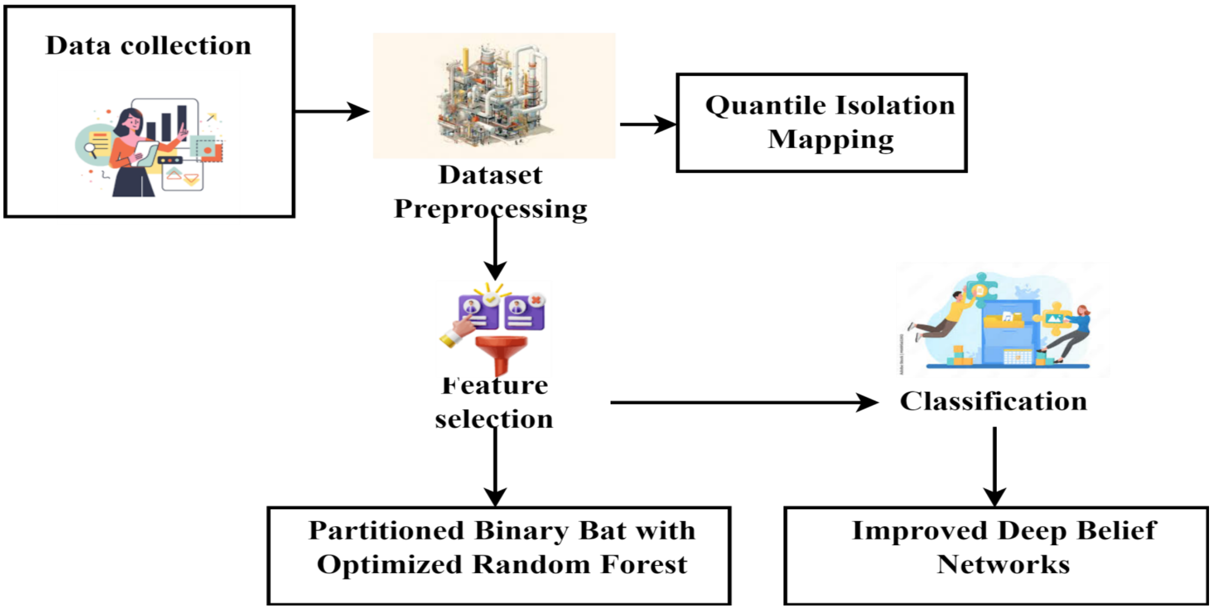

Figure 1: Proposed workflow architecture

3.2 Dataset preprocessing using Quantile Isolation Mapping

The goal of Quantile Isolation Mapping (QIM), a data preparation approach, is to effectively deal with outliers. Improving the dataset's overall robustness is the goal of this procedure, which comprises eliminating outliers and mapping data points to quantiles. By identifying and removing outliers, Quantile Isolation Mapping (QIM) boosts predictive performance, which in turn enhances the validity and precision of subsequent analyses and models.

One common post-processing step in QM techniques for climate modelling findings is the application of statistical transformations. In statistical transformations, a mathematical function is used to change the distribution functions of the modelled variables into the observed numbers. A more formal way to express this function would be as

$x^0 = f(x^m)$ ------------ (1)

3.3 Feature selection using Partitioned Binary Bat with Optimized Random Forest

One form of bat, the microbat, has the extraordinary ability to use echolocation. The most basic way that bats determine how far something is away is by producing a short burst of sound and then listening for the echo to come back. Inspired by bats' talents, a new meta-heuristic optimisation method dubbed BAT Algorithm (BA) was created. Bats use their echolocation abilities to find food using this system. A review of bat echolocation characteristics and some idealised ideas based on bat behaviour are presented below.

$Freq_i = Freq_{min} + (Freq_{max} - Freq_{min}) * \beta$ ------------ (2)

$Vl_i(t+1) = Vl_i(t) + (Pos_i(t) - G_{best}) * Freq_i$ ----------- (3)

$Pos_i(t+1) = Pos_i(t) + Vl_i(t+1)$ --------------- (4)

3.4 Classification using improved Deep Belief Networks

A state-of-the-art method for classification problems that takes use of deep learning improvements is the Improved Deep Belief Network (IDBN). Worker attrition categorization is only one example of how IDBNs outperform more conventional approaches by revealing complex data patterns and connections. Organisations can get useful insights for proactive management and decision-making with the help of IDBNs, which achieve improved accuracy in outcomes prediction via the use of feature selection and training on refined datasets.

The learnt parameters, θ1, of the network's visible and first hidden layer (RBM1), determine $p(v|\theta_1)$ using Eqns. 7 and 8. It is possible to express p(v|θ1) in this way:

$p(v|\theta_1) = \sum_h p(h|\theta_1)p(v|h,\theta_1)$ -------- (5)

Assuming RBM1 specifies $p(h|\theta_1)$, the idea behind training a DBN with a stack of RBMs is to improve $p(v)$ by replacing $p(h|\theta_1)$ with a better prior over the hidden vectors. Assuming N training instances use RBM1's hidden vectors, the "aggregated posterior" (a weighted mixture of the posterior distributions) has a higher KL divergence than the improved prior, p(v) needs improvement:

$\frac{1}{N}\sum_{v \epsilon train} p(h|v,\theta_1)$ ------------ (6)

For Gaussian mixture models, the corresponding claim is that, as training progresses, each component's updated mixing fraction should approach its average posterior probability.

Training RBM2, the network constructed from samples extracted from the aggregated posterior of RBM1, is now being considered.

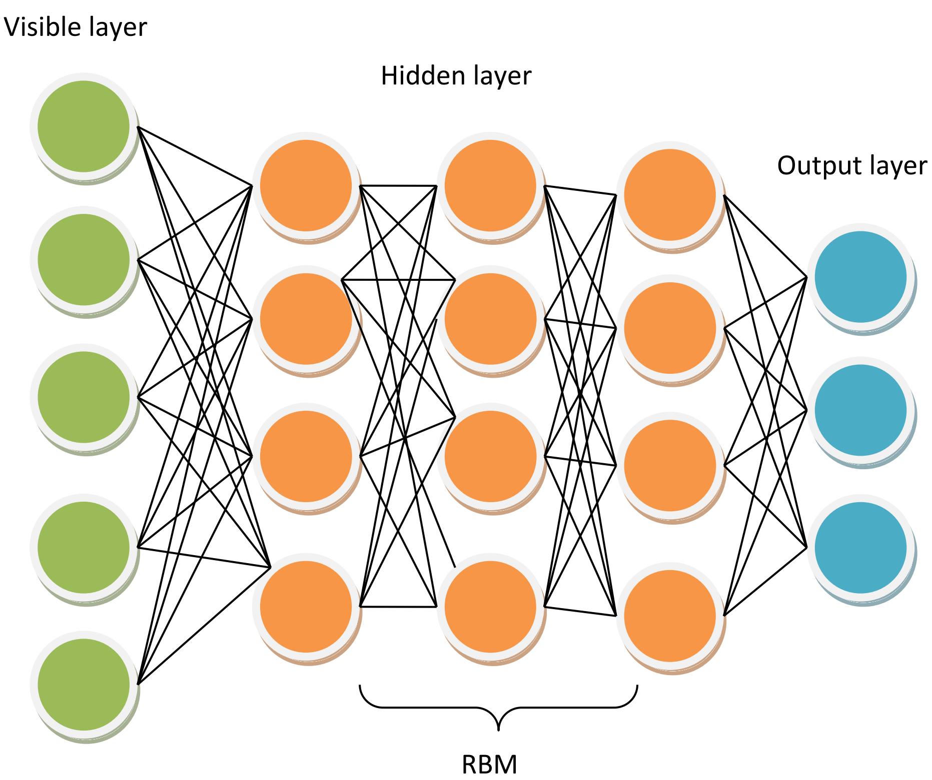

Figure 2: Deep Belief Networks

Once RBM2 has been trained, we can merge it with RBM1 to produce a model that combines the best features of directed and undirected learning. Nevertheless, as shown in, using Eqn. 7 for approximation inference on the first hidden layer improves the variational lower limit on the log likelihood of the training data with each additional layer added to the DBN, as long as the addition is done correctly.

Algorithm 1: Improved Deep Belief Networks
Input: ☐ Training dataset: $\{v(i)\}i = 1N$, where v is the visible vector. ☐ Number of hidden layers L. ☐ Learning rate α. Steps:

☐ **Initialize RBM1:**

- Initialize weights and biases for RBM1.
- Train RBM1 using Contrastive Divergence (CD) algorithm for TT epochs.

☐ **Calculate Aggregated Posterior:**

- Use RBM1 to calculate the aggregated posterior:

  1N∑v∈trainp(h|v,θ1)N1v∈train∑p(h|v,θ1)

☐ **Initialize RBM2:**

- Initialize RBM2 with visible units hh (samples from aggregated posterior) and hidden units h2h2.
- Train RBM2 using CD algorithm for TT epochs.

☐ **Combine RBM1 and RBM2:**

- Combine RBM1 and RBM2 to form a deep belief net (DBN).
- Use approximate inference for the first hidden layer:

  $p(h \mid v, \theta 1, \theta 2) \approx p(h \mid \theta 1) p(h \mid v, \theta 1, \theta 2) \approx p(h \mid \theta 1)$

☐ **Inference and Prediction:**

- Given a new input vnewvnew, perform inference through the DBN to obtain p(h|vnew,θ1,θ2)p(h|vnew,θ1,θ2).
- Classify vnewvnew as attrition-prone or stable based on the inferred probabilities.

Output:

Trained IDBN model with improved classification capabilities.

IV. Results and discussion

In this section, we delve into the results and discussion of our study on employee attrition prediction using Improved Deep Belief Networks (IDBN). We present the outcomes of our advanced approach, focusing on the model's accuracy, precision, recall, and F-measure in classifying employees as attrition-prone or stable.

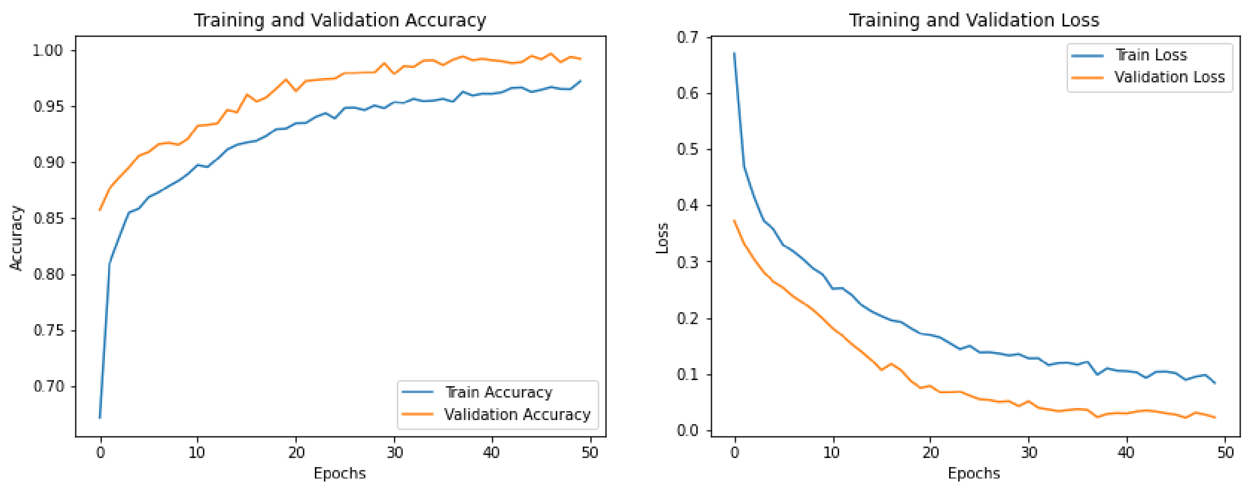

Figure 3: training and validation accuracy comparison chart

A comparison chart showing the training and validation accuracy can be seen in figure 3. Values for accuracy and loss are shown on the y-axis, while the x-axis displays epochs.

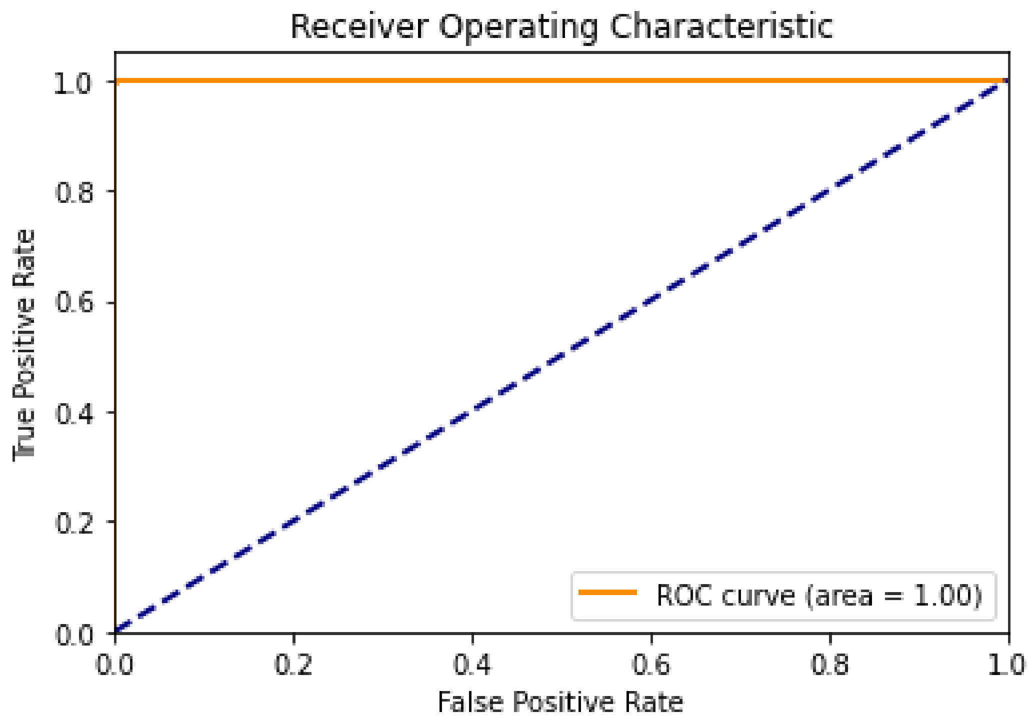

Figure 4: Receiver operating characteristic

In Figure 4, we can see the ROA graph. The false positive rate is shown on the x-axis, while the actual positive rate is shown on the y-axis.

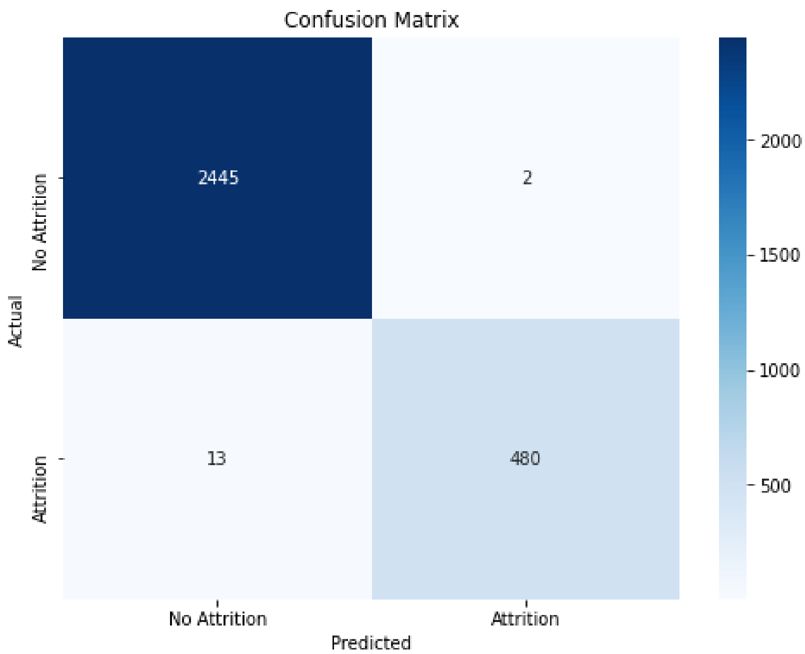

Figure 5: Confusion matrix

The figure 5 shows confusion matrix the TP is 2445 and the TN is 480 the FP is 2 and the FN is 13

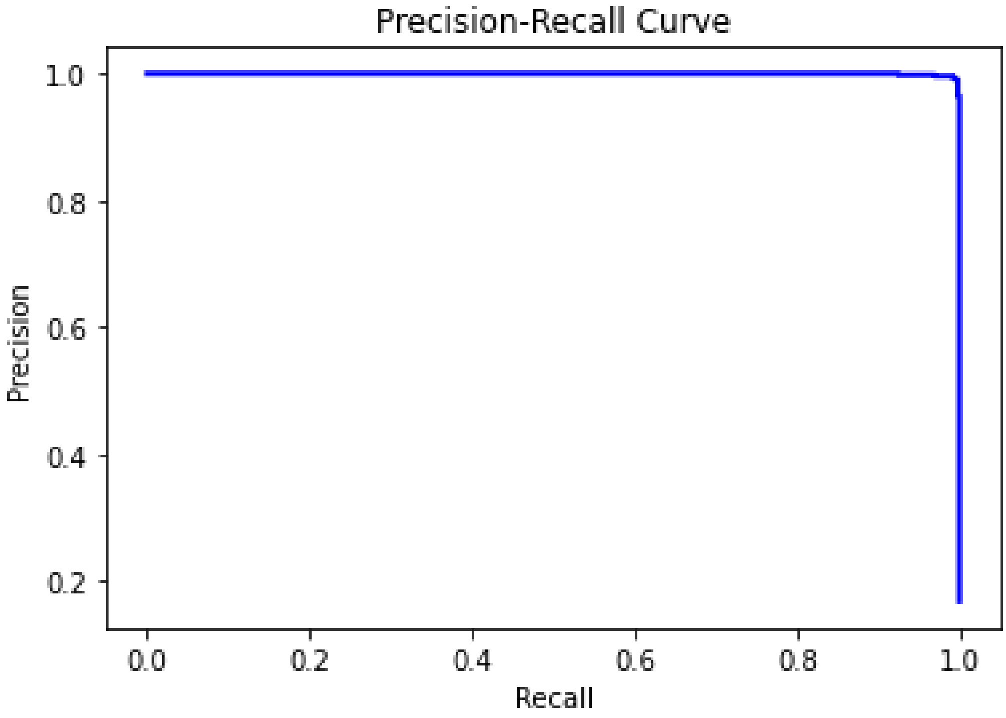

Figure 6: Precision-Recall curve

The accuracy recall curve is shown in figure 6. On one side, we have recollection, and on the other, accuracy.

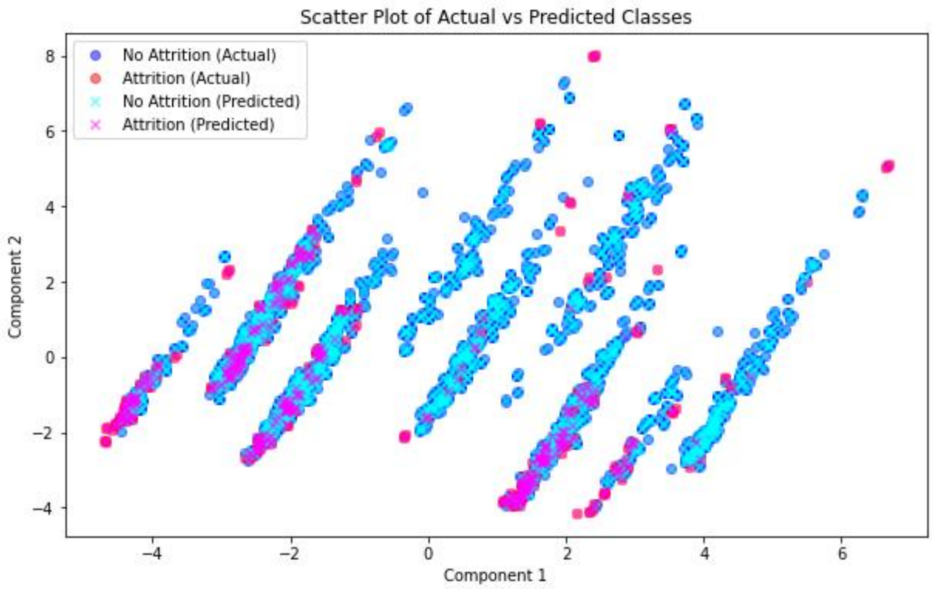

Figure 7: Scatter plot of actual vs predicted classes

A scatter plot comparing the actual and projected classes is shown in figure 7. Component 2 is shown on the y-axis while component 1 is shown on the x-axis.

Table 2: Classification metrics comparison table

Algorithms	**Accuracy**	**Precision**	**Recall**	**Fmeasure**
DT	93	93	94	93
SVM	96	94	97	96
NB	94	93	95	94
LR	96	95	96	95

CNN	97	96	97	95
DBN	98	96	97	97
Improved DBN	99.76	100	99	99

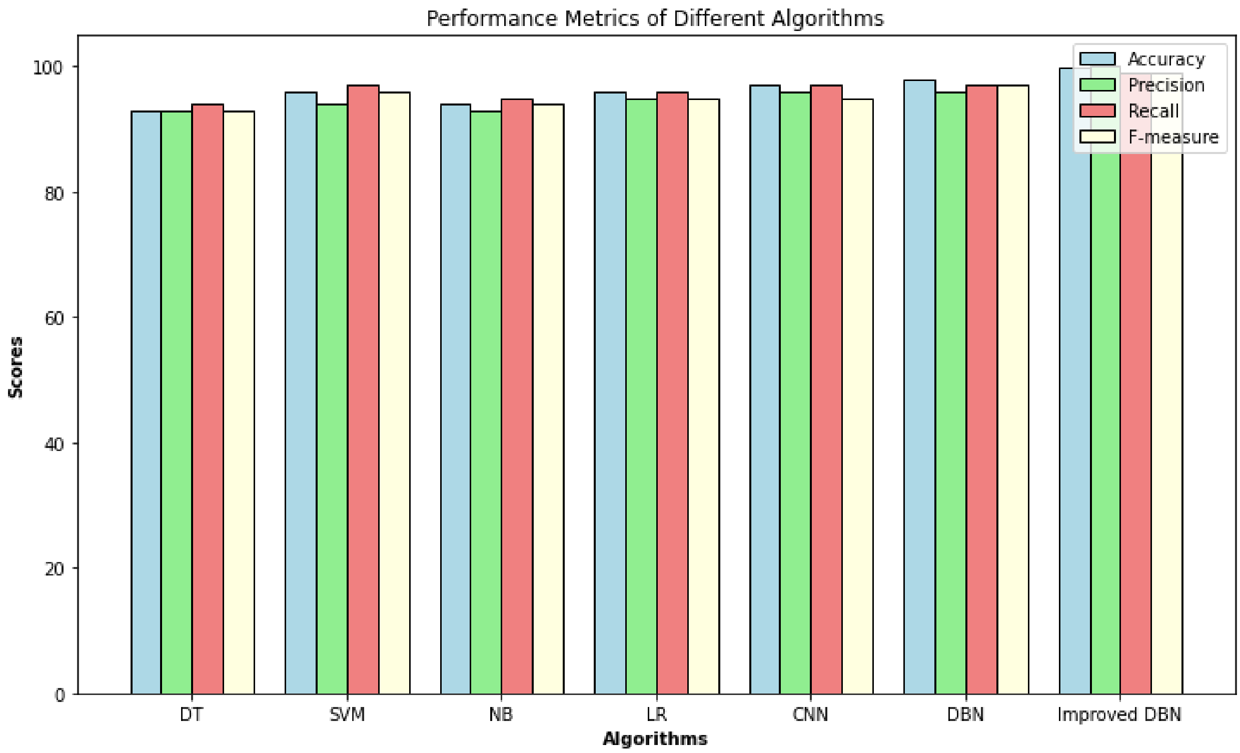

Figure 8: Classification metrics comparison chart

In order to anticipate which machine learning algorithms would be most effective in retaining employees, we examined their accuracy, precision, recall, and F-measure (see table 2 and figure 8 for details). A 93% F-measure, 94% recall, 93% precision, and 93% accuracy were all attained via the Decision Tree (DT) method. An F-measure of 96%, recall of 97%, precision of 94%, and accuracy of 96% were all shown using Support Vector Machine (SVM), indicating an improvement. A similar level of success was shown by Naive Bayes (NB), which achieved 94% accuracy, 93% precision, 95% recall, and 94% F-measure. The results obtained by Logistic Regression (LR) and Convolutional Neural Network (CNN) were quite similar; both networks achieved 96% accuracy, 95% precision, 96% recall, and 95% significance level. Further progress was shown by Deep Belief Network (DBN), which achieved 98% accuracy, 96% precision, 97% recall, and 97% F-measure. With an F-measure of 99%, a recall of 99%, a precision of 100%, and an accuracy of 99.76%, the Improved DBN algorithm outperformed all other algorithms. According to these findings, the Best algorithm for forecasting employee turnover is the Improved DBN algorithm, which produces very exact forecasts backed by strong recall and F-measure scores.

V. Conclusion

In conclusion, the utilization of Improved Deep Belief Networks (IDBN) for Employee Attrition Classification presents a significant advancement in predictive analytics for organizational management. By integrating IDBN with feature selection techniques, this research has demonstrated improved accuracy and performance in identifying attrition-prone employees. The experimental results underscore the efficacy of this approach, showcasing its potential to outperform traditional methods and provide actionable insights for proactive attrition management. Moving forward, this research opens avenues for further exploration and implementation of advanced deep learning techniques in workforce analytics. By leveraging the full potential of available data and harnessing the power of IDBN, organizations can enhance their ability to predict and address employee attrition, ultimately leading to improved productivity, morale, and cost management. The Improved DBN algorithm achieved the highest scores across all metrics, with an accuracy of 99.76%, precision of 100%, recall of 99%, and an F-measure of 99%. This study contributes valuable knowledge and methodologies for organizations seeking to optimize workforce stability and drive sustainable business outcomes.

References

[1] Alshiddy, M. S., & Aljaber, B. N. (2023). Employee Attrition Prediction using Nested Ensemble Learning Techniques. *International Journal of Advanced Computer Science and Applications*, *14*(7).

[2] Doss, S. (2022). Analytic Approach of Predicting Employee Attrition Using Data Science Techniques.

[3] El-Rayes, N., Fang, M., Smith, M., & Taylor, S. M. (2020). Predicting employee attrition using tree-based models. *International Journal of Organizational Analysis*, *28*(6), 1273-1291.

[4] Engl, F., & Herrmann, F. (2023). A Machine Learning based Approach on Employee Attrition Prediction with an Emphasize on predicting Leaving Reasons. *Anwendungen und Konzepte der Wirtschaftsinformatik*, (18), 11-11.

[5] Jain, D. (2017). *Evaluation of employee attrition by effective feature selection using hybrid model of ensemble methods* (Doctoral dissertation, Dublin, National College of Ireland).

[6] Jain, P. K., Jain, M., & Pamula, R. (2020). *Explaining and predicting employees' attrition: a machine learning approach. SN Applied Sciences, 2(4).* doi:10.1007/s42452-020-2519-4

[7] Kumari, S. (2023). *Leveraging Machine Learning to Predict Employee Attrition: India* (Doctoral dissertation, Dublin, National College of Ireland).

[8] Mohbey, K. K. (2020). Employee's attrition prediction using machine learning approaches. In *Machine Learning and Deep Learning in Real-Time Applications* (pp. 121-128). IGI Global.

[9] Norrman, F. (2020). Predicting employee attrition with machine learning on an individual level, and the effects it could have on an organization.

[10] Özdemir, F. (2020). *Recommender System For Employee Attrition Prediction And Movie Suggestion* (Master's thesis, Abdullah Gül Üniversitesi, Fen Bilimleri Enstitüsü).

[11] Raj, A. (2022). *Deciphering the Augmentation of Classification Models in Predicting Employee Attrition* (Doctoral dissertation, Dublin, National College of Ireland).

[12] Ray, A. N., & Sanyal, J. (2019). *Machine Learning Based Attrition Prediction. 2019 Global Conference for Advancement in Technology (GCAT).* doi:10.1109/gcat47503.2019.8978285

[13] Sethy, A., & Raut, A. K. (2020). Employee attrition rate prediction using machine learning approach. *Turkish Journal of Physiotherapy and Rehabilitation*, *32*(3).

[14] Soni, U., Singh, N., Swami, Y., & Deshwal, P. (2018, September). A comparison study between ANN and ANFIS for the prediction of employee turnover in an organization. In *2018 International Conference on Computing, Power and Communication Technologies (GUCON)* (pp. 203-206). IEEE.

[15] Taylor, S., El-Rayes, N., & Smith, M. (2020). An Explicative and Predictive Study of Employee Attrition Using Tree-based Models.

[16] Veerkhare, R., Shah, P., Sidhpura, J., & Dhage, S. (2022, March). HRescue: A Modern ML Approach for Employee Attrition Prediction. In *International Conference on Machine Learning and Big Data Analytics* (pp. 279-293). Cham: Springer International Publishing.

[17] Yadav, S., Jain, A., & Singh, D. (2018, December). Early prediction of employee attrition using data mining techniques. In *2018 IEEE 8th international advance computing conference (IACC)* (pp. 349-354). IEEE.

[18] Garg, U., Gupta, N., Manchanda, M., & Purohit, K. (2024, April). Classification and Prediction of Employee Attrition Rate using Machine Learning Classifiers. In 2024 International Conference on Inventive Computation Technologies (ICICT) (pp. 608-613). IEEE.