

An Artificial Neural Network and Infographic Based Approach for Predicting Employee Attrition

***Mamillapali Dianasaroj**

School of Computer Science and Engineering, Sandip University Nashik

*dianasarojcc@gmail.com

Pravin Gundalwar

School of Computer Science and Engineering, Sandip University Nashik

pravin.gundalwar@sandipuniversity.edu.in

Abstract - Making decisions is an essential managerial skill and can be the most important step in the planning process. Employee attrition is seen as a well-known issue that the administration needs to address if it wants to keep highly qualified staff. It's interesting to note that using artificial intelligence as a powerful tool to foresee such a problem is commonly done. Utilizing the ANN deep learning model is the proposed study. To improve the prediction of employee attrition, artificial neural networks also conduct comparison analysis with earlier research models and use a variety of preprocessing methods. The proposed model's maximum Training Accuracy or Validation Accuracy were 98.25 and 89.88 respectively, in comparison to earlier research models like SVM, LSTM, & GRU. The proposed model ANN produced the lowest Validation Loss of 0.0674 with the Lowest Training Loss of 0.3351, respectively.

Keywords- Infographics, Data Science, Employee Attrition, Deep learning, Artificial Neural Network, SMOTE, One Hot encoding.

Introduction

Attrition is the term used to describe when a worker leaves a company without being replaced. In other words, a new employee won't be hired to fill the position the person held. Since vacant positions cannot be filled, attrition results in layoffs and, ultimately, a smaller workforce. Employment turnover can be caused by voluntary or involuntary factors. The attrition rate is the proportion of employees in a company who leave each year. In its broadest sense, attrition refers to the gradual weakening or reduction of an object, as when friction wears down stones and causes them to become smaller. When staff members quit an organization for unforeseen or unavoidable reasons, attrition happens[1]. Attrition includes situations such as dismissal, resignation, retirement (both voluntary and planned), organizational changes, chronic illness, and job losses. When employees leave a company at a faster rate than replacements can be identified, attrition takes place. When positions go unfilled for too long, they are either eliminated or not replaced at all. Sometimes a corporation can't do much to reduce staff turnover due to a shortage of competent candidates. Attrition is the term used to describe the steady exodus of members over time. The problem will require long-term, strategic solutions from business and HR executives[2]. However,

turnover is only a short-term issue. A typical short-term strategy for firms to deal with resource or talent shortages is to hire more staff. Employee attrition accounts for all lost employment and long-term openings, whereas turnover solely considers temporary roles[3]. Businesses all across the world struggle greatly with a lack of accessible labor. The year 2021 has been known in the business world as "The Great Resignation." In search of better work-life balance, healthier working environments, and general happiness and well-being, many people quit their jobs or looked for new ones. Business executives anticipated a return to normalcy in 2022 as 2021 concluded, particularly in light of the relaxation of movement restrictions. They prayed that the workers would be able to return to their jobs and adjust to the new situation[4]. In terms of the workplace of the future, though, this hasn't been without its share of challenges. For enterprises, there will soon be a significant administrative adjustment. As a result, HR managers and staff have been inundated with requests for hybrid schedules, updated policies, and higher compensation and benefits. During these times of change, it is critical to control attrition[5]. A company's turnover rate is a crucial sign of how employee-friendly it is and where it may make improvements. High turnover rates may be attributed to a variety of factors, including inadequate compensation and benefits, an antiquated, high-pressure workplace environment, or a lack of opportunities for advancement. Whatever the reason, a high turnover rate and the time and effort needed to fill these positions are both expensive problems for any business. Effective recruiting requires time and money for tasks including hiring, onboarding, and training. According to a Gallup survey, the expense of replacing a full-time employee might range between 50% and 200% of their annual salary. The time it takes new hires to become familiar with aspects of the workplace culture, and form relationships with coworkers and procedures must be taken into consideration. When an employee departs, they provide more than just their knowledge[6]. People frequently take important contacts, knowledge, and personal networks with them when they leave an organization. Businesses should keep track of their worker turnover rate to save money, time, and energy[7][8].

By employing data as fuel, data science is the force driving numerous technologies toward automation in the modern world. The study of data is what is meant by data science. Students go over a beginner's Data Science lecture as they begin to visualize and analyze data. to analyze and present a large amount of data, use a variety of statistical approaches. Statistics is the foundation of data science, deep learning, artificial intelligence, and machine learning. Layered architecture-based techniques are developed in the machine learning subfield known as deep learning. The organic human brain network serves as a model for the layered architecture of deep learning (deep neural network). Since the system uses a thick neural network, it is referred to as a "deep neural network." This deep learning network executes the full machine learning process.

In recent years, deep learning algorithms have become more and more useful for predicting employee attrition in the field of computer science. These systems offer estimations based on previously collected data about the employee, like age, experience, education, last promotion, and so forth. The results of the prediction have informed the HR department about staff

attrition. To fill the position of the employee who plans to leave soon, the HR department has already prepared to hire replacement employees[9][10].

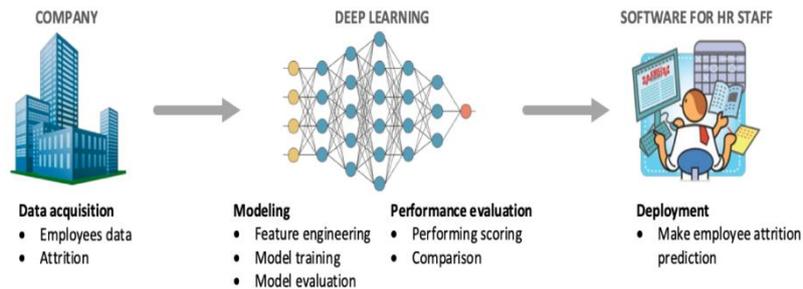


Fig. 1 Employee attrition using deep learning

Literature Review

Mduma 2023 et. al [11] to improve minority class prediction accuracy while maintaining a reasonable overall classification performance. tested several lesser-known classification methods, including Random Over Sampling, Random Under Sampling, Synthetic Minority Over Sampling, SMOTE through Edited Nearest Neighbour, as well as SMOTE with Tomek connections, in addition to the well-known classification methods of Logistic Regression, Random Forest, and Multi-Layer Perceptron. Using Tanzanian and Indian datasets that are freely available, the efficacy of balancing techniques and prediction models was evaluated. The results show that SMOTE with Edited Nearest Neighbour performed the best in terms of classification on the 10-fold holdout sample. Logistic Regression was successfully used to determine the greatest number of students who dropped out when the confusion matrix was used as the assessment matrix (57348 for the Uwezo dataset and 13430 for the India dataset). Finding at-risk students and reducing attrition rates are both attainable using these methods.

Pratibha 2022 et. al [12] impact turnover rates significantly. In this study, the likelihood of each new hire departing was predicted using a variety of classification techniques, including ridge classification, decision trees, and random forests. Each one of these supervised machine-learning techniques is assessed utilizing a thorough evaluation process. The survey's findings can help human resource managers anticipate which employees will quit the organization and for what reasons, giving them more time to either implement a retention strategy or find a replacement.

Atef 2022 et. al [13] Using k-nearest neighbors with random forests as machine learning techniques. The models benefit from a dataset IBM created on staff turnover. The most crucial characteristics that are taken into account when hiring new employees and which can cause turnover are included in the dataset that was used. These characteristics include gender, salary, age, and distance from home. Regarding precision and accuracy, the KNN-based model performed better compared to the RF-based model in terms of -score, specificity (SP), & false-positive rate (FPR). The normal percentage of employees who intend to quit their jobs is predicted by the algorithms. Human resource managers can use the models to help them make wise decisions about whether a candidate employee is likely to stay in the job or

leave it in light of the relevant information that has been provided about the candidate employee.

Aggarwal 2022 et. al [14] Due to the harm it causes to staff productivity and the capacity to meet the company's vision and goals on time, all organizations view it as a major issue. Businesses are adopting machine learning techniques more and more to predict the possibility of workforce turnover to address this problem. The main objective of these methods is to produce goods that profit from the analysis of employee data and the precision of predictive models. Businesses can make prompt judgments regarding employee retention or termination by using credible data. Projecting a worker's future productivity for the company using the current method, The system for tracking human resources has a lot of errors. utilizing cutting-edge machine learning algorithm models, we anticipate employee attrition in this study utilizing a large, existing data set. Employee attrition is the intention of a worker to leave or stay with the company. Making informed decisions is no longer facilitated by the business models currently in use.

Barramuño 2022 et. al [15] To supply the information, a platform at the University of Chile was used to access the Financial Academic Administration and Management System. College student attrition rates were examined using five quantitative and eleven qualitative criteria. 23 supervised machine learning classifiers based on this database were evaluated. According to the findings, there was an attrition rate of about 23.58% for male students and 17.39% for female students. The classifiers' mean accuracy increased when more variables were used during training. The best level of accuracy (86.3%) was attained with the "Subspace KNN" approach. The "RUSboosted trees" classifier outperforms all other investigated methods in terms of sensitivity (78%) and specificity (86%). It has practical implications since this prediction technology can be used to lower student dropout rates and increase retention rates among advanced students in academic programs. The study's unique prediction model for student attrition in upper-year courses works for unbalanced databases with fewer attrition students, underscoring the study's uniqueness and boosting its value.

Table.1 Summary of Literature

Author/ year	Method/ Model	Parameters	References
Soner/2022	SVM SMOTE	Accuracy= 89%, recall= 92%, precision= 94% and f1 Score= 93%	[16]
Shete/2021	SVM, KNN	--	[17]
Korichi/2021	--	Accuracy= 0.83%	[18]
Mansor/2021	SVM, ANN	Accuracy= 88.87%	[19]
G/ 2021	machine learning methods and	F1 SCORE= 0.283	[20]

Proposed Methodology

This discussion centers on how to recognize human activity. The first step is to obtain data from the IBM Hr. To get rid of any extraneous values that may have been there, the data is then preprocessed utilizing data filtering, data removal, and discovering null values. Next, One Hot encoding and SMOTE are finished, followed by ANN modeling. Next, the forecast is evaluated. The proposed flowchart is shown in Fig. 2.

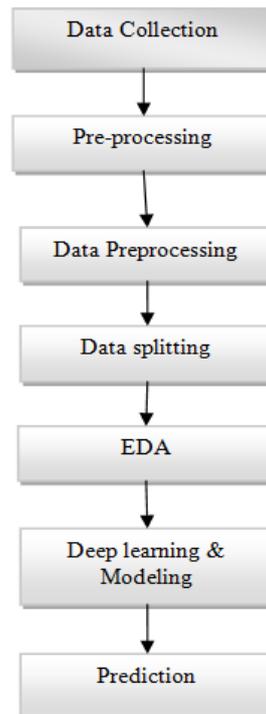


Fig.2 Proposed Flowchart

A. Data Collection

Data collected from <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset> A dataset is a group of papers used for research and analysis. You can add importers & crawlers to the dataset to import your data. You can add your data to collections for analysis after it has been loaded into the dataset. Take a close look at the elements that affect employee turnover and give careful consideration to crucial inquiries like "Show me a breakdown of commute time by job role and attrition" or "Compare average monthly income on education with attrition." The data scientists at IBM produced this hypothetical data set. The phrase "Below College" uses the words education, college, bachelor, master, and doctor. feeling content with the surrounding.

B. Data Preprocessing

Complete three different steps within the entire data preparation activity. These specific tasks include One Hot and SMOTE encoding. processing data to identify seasonality in time series data, for instance, or performing exploratory research for data visualization. After that, the

data must be transformed into a three-dimensional NumPy array so that a recurrent neural network may be trained. To analyze and filter data, one can use none, null, and fill null values included in the Panda library. filling in the null values with the median method included in the Pandas data frame. This approach makes use of a statistical technique to determine the average value (the middle) of a set of data. After the data have been sorted ascendingly, the median is then calculated. Pictures of data samples are shown in the image below. Equation 1 displays one Hot encoding.

$$\sum_{i=1}^M (y_i - \sum_{j=0}^p w_j * x_{ij})^2 + \lambda \sum_{j=0}^p w_j^2 \quad (1)$$

C. Data Splitting

The data will be split into training and testing phases having an 80:20 ratio before the classification model is developed. Training data, not testing or validation data, are utilized when the deep learning tool has been trained to recognize patterns or satisfy predetermined requirements.

D. Deep learning & Modeling

Artificial neural networks are the core of the "deep learning" subfield of machine learning. It can find complex relationships and patterns in data. Not everything in deep learning requires explicit programming. Its application has grown recently as a result of the availability of huge datasets and advancements in processing speed. because it is based on deep neural networks (DNNs), also referred to as artificial neural networks (ANNs). The shape and operation of these neural networks, which are designed to learn from a large amount of data, are inspired by the biological neurons in the human brain.

- Artificial Neural Network (ANN)

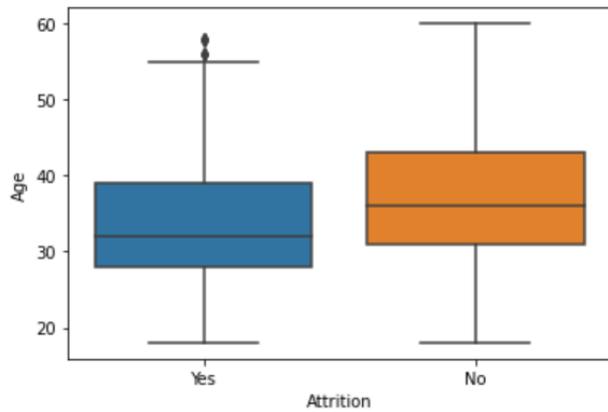
A type of computing network called artificial neural networks (ANNs) is biologically based. The multilayer perceptron (MLP), which uses backpropagation learning methods, is one of the several varieties of ANNs being investigated. The most popular ANNs, known as MLPs, are constructed using supervised learning techniques and feature three layers: input, hidden, and output. We talk about the structure, algorithm, data pretreatment, overfitting, and sensitivity analysis of MLPs, among other topics. The benefits and drawbacks of MLPs are also discussed, and their usage in ecological modeling is advised. The last example shows how MLP can be used in ecological models practically. Equation 2 shows the ANN equation.

$$W_x = W_x - a \left(\frac{\partial Error}{\partial W_x} \right) \quad (2)$$

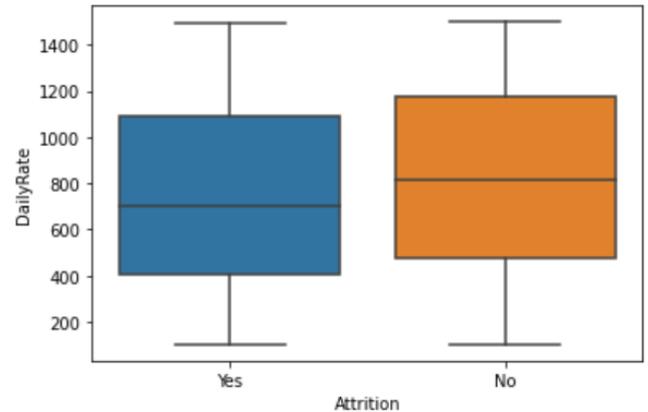
E. Exploratory data analysis (EDA)

EDA is largely used to examine what data might disclose, in addition to the formal modeling and hypothesis testing effort. Additionally, it offers a deeper understanding of the factors used to acquire the data and how they relate to one another. You can also use it to check the suitability of the statistical techniques you were thinking of using for your data analysis. If a

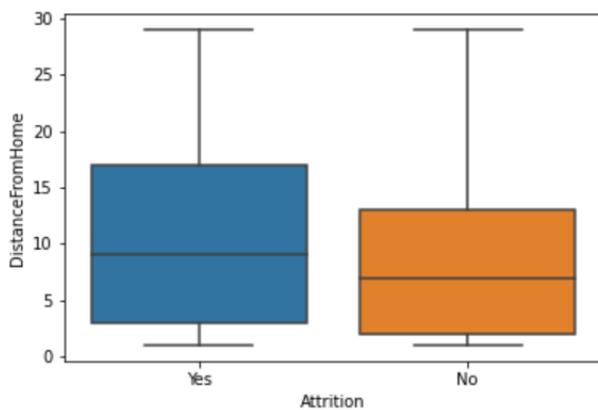
data point considerably deviates from both its immediate surroundings and the rest of the dataset by being observably higher or lower, it is referred to as an outlier.



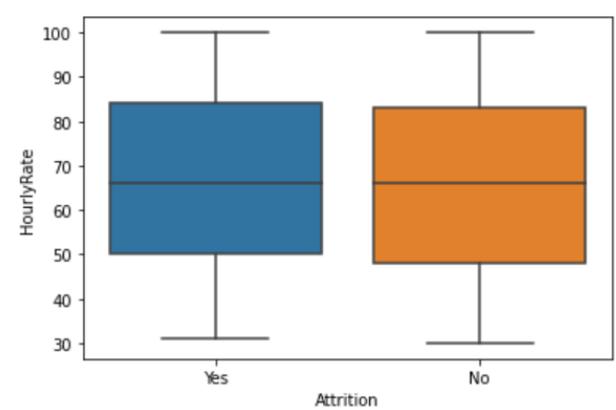
(a)



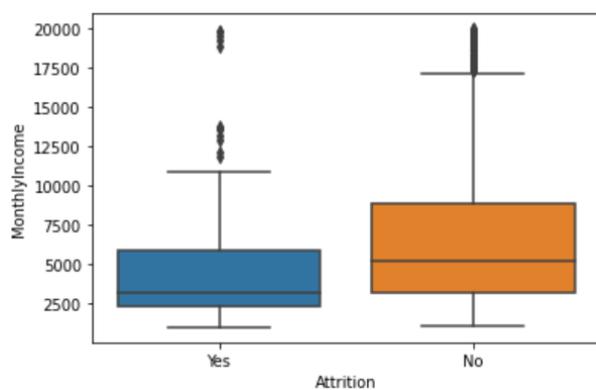
(b)



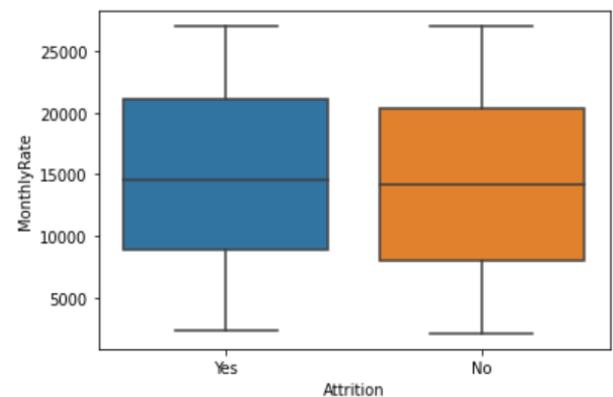
(c)



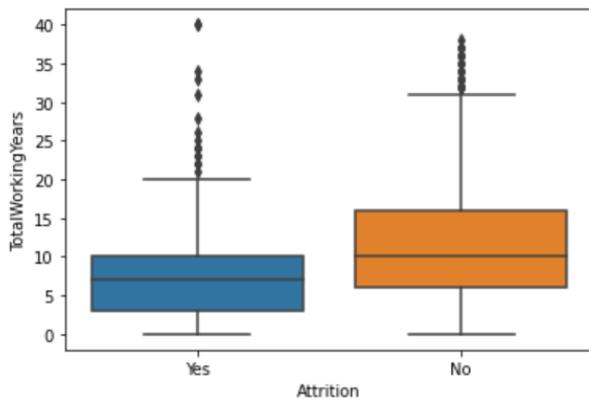
(d)



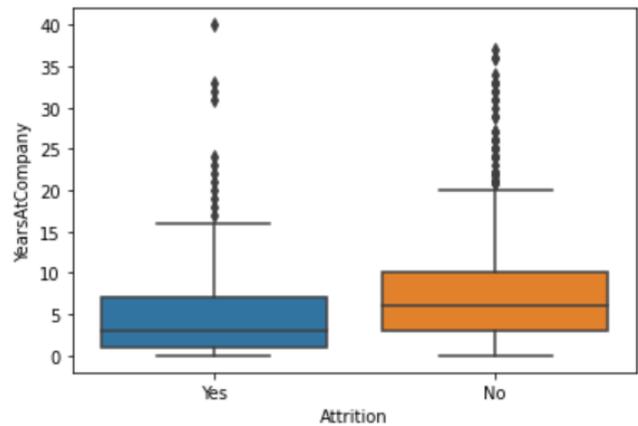
(e)



(f)



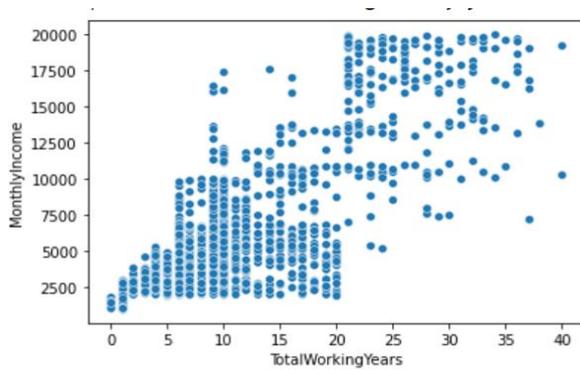
(g)



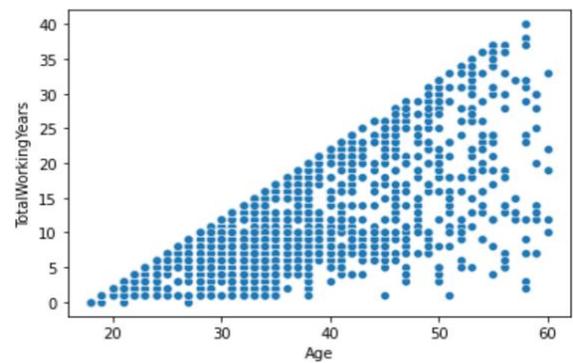
(h)

Fig.3 Boxplots of Data Variables

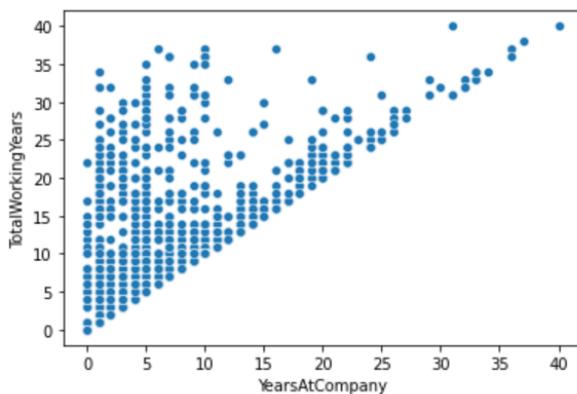
Fig. 3 displays data visualization using boxplots for the data variables ((a)Age, (b)Attrition,(c) distance from home,(d) Hourly rate,(e) Monthly Income,(f) Monthly rate, (g) Total working years, (h) Years at company .where orange indicates no and blue indicates yes.



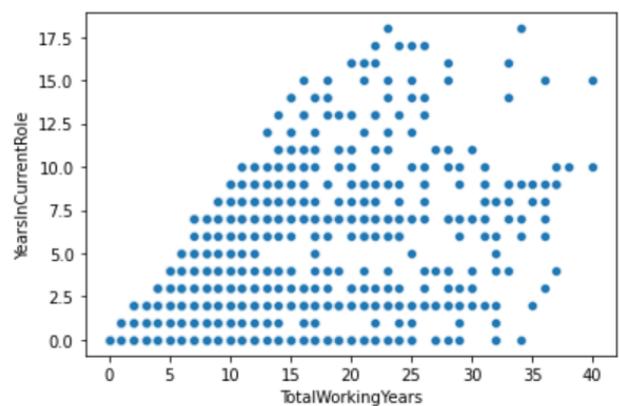
(a)



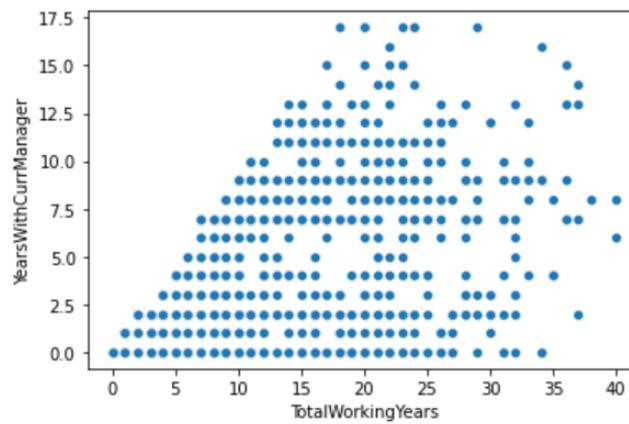
(b)



(c)



(d)



(e)

Fig .4 Scatter plots of Data variables.

Fig. 4 displays scatter plots for data variables such as (a) monthly income and the total working year, (b) age and the total working year, (c) total working years and years at the company, (d) total working years with years in a current role, as well as (e) total working years and years with current management.

Result & Discussion

An explanation of the deep learning models' findings is provided in this section. In this work, an ANN-based suggested model is used to quantify the accuracy and loss for the prediction of employee attrition.

1) Accuracy

One of the metrics used to assess classification models is accuracy. The percentage of occurrences that were accurately predicted by our model is called accuracy. Accuracy is defined as follows in a formal sense.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (9)$$

2) Loss

Since neural networks are trained using stochastic gradient descent, you must choose a loss function while building and configuring your model. When training a neural network, it might be challenging to choose the loss function to employ or even understand what a loss function is and what it does.

$$Loss = -\frac{1}{m} \sum_{i=1}^m y_i \cdot \log(\hat{y}_i) \quad (10)$$

Table.1 Performance Evaluation of Models

Model	Training Acc	Training Loss	Validation Acc	Validation Loss
Proposed ANN	98.25	0.0674	89.88	0.3351
SVM	88.98	0.1244	66.91	0.8611
LSTM	93.44	0.8443	83.76	0.5849
GRU	95.32	0.7644	69.31	0.5632

The maximum Training Accuracy and Validation Accuracy achieved by the Proposed model ANN at 98.25 or 89.88 respectively are presented in Table.2's performance evaluation of the Proposed Model ANN and comparison to prior research models like SVM, LSTM, and GRU. The lowest Validation Loss of 0.0674 and Lowest Training Loss of 0.3351 were obtained using the suggested model ANN, respectively. Proposed model ANN improved by 10% compared to SVM, 5% compared to LSTM, and 3% compared to GRU.

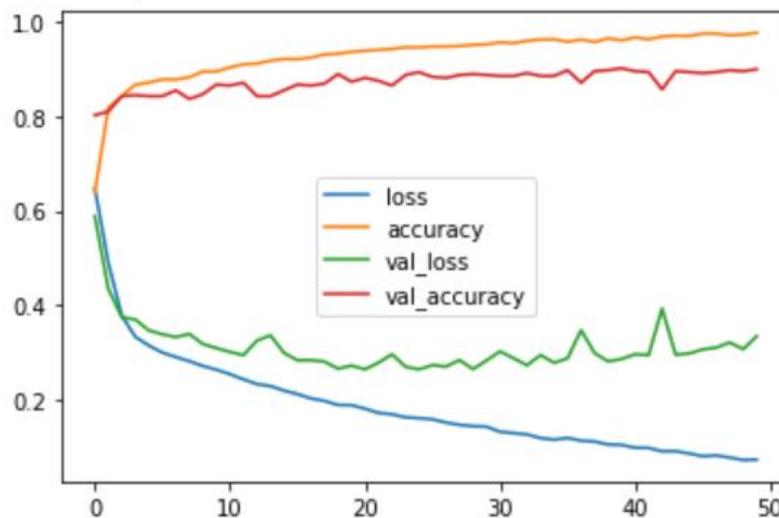


Fig.5 Accuracy and Loss Graph

The accuracy, loss, accuracy, and validation accuracy & validation loss of the proposed model ANN are shown in Fig. 5, where the blue line denotes loss, the orange color denotes accuracy, the green color denotes validation loss, and the red color denotes validation accuracy.

conclusion

The decision-making process itself may be the most crucial step in the planning process and a crucial managerial skill. It is widely acknowledged as the administration must address the well-known problem of employee turnover if it is to retain highly competent workers. It's astounding to observe how frequently artificial intelligence is employed as a potent tool for

anticipating such a predicament. The suggested inquiry employs the ANN deep learning model. Artificial neural networks use a variety of preprocessing techniques and comparison analyses with past research models to enhance the prediction of employee attrition. Compared to preceding research models like SVM, LSTM, & GRU, the proposed model's maximum Training Accuracy and Validation Accuracy were 98.25 and 89.88, respectively. The lowest Validation Loss of 0.0674 and Lowest Training Loss of 0.3351 were both achieved by the suggested ANN model. Compared to SVM, LSTM, and GRU, The proposed model ANN outperformed SVM, LSTM, and GRU by 10%, 5%, and 3%, respectively.

References

- [1] dkk 2018) richard oliver (dalam Zeithml., “濟無No Title No Title No Title,” *Angew. Chemie Int. Ed.* 6(11), 951–952., no. May, pp. 2013–2015, 2021.
- [2] S. Al-Darraj, D. G. Honi, F. Fallucchi, A. I. Abdulsada, R. Giuliano, and H. A. Abdulmalik, “Employee attrition prediction using deep neural networks,” *Computers*, vol. 10, no. 11, pp. 1–11, 2021, doi: 10.3390/computers10110141.
- [3] İ. ERSÖZ KAYA and O. KORKMAZ, “Machine Learning Approach for Predicting Employee Attrition and Factors Leading to Attrition,” *Çukurova Üniversitesi Mühendislik Fakültesi Derg.*, vol. 36, no. December, pp. 913–928, 2021, doi: 10.21605/cukurovaumfd.1040487.
- [4] I. Ballal, S. Kavathekar, S. Janwe, P. Shete, and N. Bhirud, “People Leaving the Job-An Approach for Prediction Using Machine Learning,” *Int. J. Res. Anal. Rev.*, vol. 7, no. 1, pp. 8–10, 2020.
- [5] F. Norrman, “DEGREE PROJECT IN THE FIELD OF TECHNOLOGY Predicting employee attrition with machine learning on an individual level , and the effects it could have on an organization,” 2020.
- [6] M. Liu, “Using Machine Learning Models to Predict Attrition in a Survey Panel,” *Big Data Meets Surv. Sci. A Collect. Innov. Methods*, pp. 415–433, 2020, doi: 10.1002/9781118976357.ch14.
- [7] P. K. Jain, M. Jain, and R. Pamula, “Explaining and predicting employees’ attrition: a machine learning approach,” *SN Appl. Sci.*, vol. 2, no. 4, pp. 1–11, 2020, doi: 10.1007/s42452-020-2519-4.
- [8] N. Sabbineni, “Understanding employee attrition using explainable AI,” no. May, 2020.
- [9] F. Fallucchi, M. Coladangelo, R. Giuliano, and E. W. De Luca, “Predicting employee attrition using machine learning techniques,” *Computers*, vol. 9, no. 4, pp. 1–17, 2020, doi: 10.3390/computers9040086.
- [10] C. Wijnen, “Predicting employee attrition,” no. May, pp. 1–62, 2019.
- [11] N. Mduma, “Data Balancing Techniques for Predicting Student Dropout Using Machine Learning,” *Data*, vol. 8, no. 3, p. 49, 2023, doi: 10.3390/data8030049.
- [12] G. Pratibha and D. N. Hegde, “Analytical Survey on Prediction of Employee Attrition Non parametric tuning Algorithms,” vol. 10, no. 6, pp. 818–826, 2022.
- [13] M. Atef, D. S. Elzanfaly, and S. Ouf, “Early Prediction of Employee Turnover Using Machine Learning Algorithms 135 Original Scientific Paper,” *Int. J. Electr. Comput.*

Eng. Syst., pp. 135–144, 2022.

- [14] S. Aggarwal, M. Singh, S. Chauhan, M. Sharma, and D. Jain, “Employee Attrition Prediction Using Machine Learning Comparative Study,” *Smart Innov. Syst. Technol.*, vol. 265, no. 9, pp. 453–466, 2022, doi: 10.1007/978-981-16-6482-3_45.
- [15] M. Barramuño, C. Meza-Narváez, and G. Gálvez-García, “Prediction of student attrition risk using machine learning,” *J. Appl. Res. High. Educ.*, vol. 14, no. 3, pp. 974–986, 2022, doi: 10.1108/JARHE-02-2021-0073.
- [16] S. Soner, A. A. Hussain, and R. Khatri, “Predictive Deep Learning approach of employee attrition for imbalance datasets using SVM SMOTE algorithm with Bias Initializer,” pp. 1–19, 2022.
- [17] M. B. Shete, “Prediction of Employee Attrition Using Machine Learning Approach,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 9, no. 8, pp. 2479–2483, 2021, doi: 10.22214/ijraset.2021.37796.
- [18] A. R. Korichi, H. Kheddouci, and D. J. West, “Predicting employee attrition with a more effective use of historical events,” *ESANN 2021 Proc. - 29th Eur. Symp. Artif. Neural Networks, Comput. Intell. Mach. Learn.*, pp. 53–58, 2021, doi: 10.14428/esann/2021.ES2021-110.
- [19] 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.
- [20] G. Klop, “Predicting Employee Turnover and Reducing Turnover Costs Using Machine Learning Techniques,” *Ind. Eng. Innov. Sci.*, p. 98, 2021.