

Research Article

An Enhanced Deep Neural Network for Predicting Workplace Absenteeism

**Syed Atif Ali Shah,¹ Irfan Uddin²,³ Furqan Aziz³,⁴ Shafiq Ahmad⁴,
Mahmoud Ahmad Al-Khasawneh,⁵ and Mohamed Sharaf⁶**

¹Faculty of Engineering and Information Technology, Northern University, Nowshehra, Pakistan

²Institute of Computing, Kohat University of Science and Technology, Kohat, Pakistan

³Center for Excellence in IT, IMSciences, Peshawar, Pakistan

⁴King Saud University, College of Engineering, Department of Industrial Engineering, Riyadh 11421, Saudi Arabia

⁵Faculty of Computer & Information Technology, Al-Madinah International University, Kuala Lumpur, Malaysia

⁶King Saud University, College of Engineering, Department of Industrial Engineering, Riyadh 11421, Saudi Arabia

Correspondence should be addressed to Irfan Uddin; irfanuddin@kust.edu.pk

Received 28 October 2019; Revised 12 January 2020; Accepted 18 January 2020; Published 19 February 2020

Academic Editor: Daniela Paolotti

Copyright © 2020 Syed Atif Ali Shah et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Organizations can grow, succeed, and sustain if their employees are committed. The main assets of an organization are those employees who are giving it a required number of hours per month, in other words, those employees who are punctual towards their attendance. Absenteeism from work is a multibillion-dollar problem, and it costs money and decreases revenue. At the time of hiring an employee, organizations do not have an objective mechanism to predict whether an employee will be punctual towards attendance or will be habitually absent. For some organizations, it can be very difficult to deal with those employees who are not punctual, as firing may be either not possible or it may have a huge cost to the organization. In this paper, we propose Neural Networks and Deep Learning algorithms that can predict the behavior of employees towards punctuality at workplace. The efficacy of the proposed method is tested with traditional machine learning techniques, and the results indicate 90.6% performance in Deep Neural Network as compared to 73.3% performance in a single-layer Neural Network and 82% performance in Decision Tree, SVM, and Random Forest. The proposed model will provide a useful mechanism to organizations that are interested to know the behavior of employees at the time of hiring and can reduce the cost of paying to inefficient or habitually absent employees. This paper is a first study of its kind to analyze the patterns of absenteeism in employees using deep learning algorithms and helps the organization to further improve the quality of life of employees and hence reduce absenteeism.

1. Introduction

The growth and success of an organization depend on its employees. Therefore, it is really important that employees are punctual towards their attendance and work for the number of hours defined by the employer organizations. Generally, the number of working hours in an organization is eight hours per day or 240 hours per month. Organizations prefer that employees are present for the maximum number of hours. However, there can be unavoidable circumstances, and hence, different types of leaves are granted

in most of the cases. In different countries, employees are allowed to take a certain amount of leaves per month, and employers expect employees to use their due right. But those who are habitually absent are the real reason why organization's productivity or revenue decreases.

Absenteeism at work can be defined as a habitual pattern of absence from a duty or obligation. Generally, absenteeism is assumed as a major indicator of poor performance. Absenteeism happens when employees are habitually late or engage in activities that are not directly or indirectly related to their work, e.g., long coffee breaks, overextended breaks,

excessive personal times, internet time, and unnecessary socialization. At work places, different strategies are used to enforce effective time utilization [1]. For instance, some organizations enforce a check-in and check-out time and may deploy different software or biometric devices to detect absenteeism at work. The behavior of employees towards absenteeism cannot be straightforward to detect and can come in numerous shapes and several stages of severity [2–4]. For instance, an employee may be sitting in his/her office for eight hours a day but may be involved in lengthy phone calls, social networking sites, games, etc. It is very important for the organization to predict the behavior of employees towards punctuality at work at an early stage, such as before calling them for interview for a vacant position. This can avoid the cost of conducting interview, hiring process, and even the hassle of dealing with people who are habitually absent and are affecting the organizational work environment.

Neural Networks (NNs) are around for many years and have been used extensively in solving different problems [5]. In the last few years, Neural Network, particularly Deep Neural Networks (DNNs), is becoming extremely popular [6, 7] and is achieving better performance compared to traditional machine learning algorithms (such as logistic regression, Decision Tree, and SVM). In this paper, we present Neural Network and Deep Neural Network that can predict the behaviour of employees towards punctuality in attendance. Organizations can deploy these models to predict the behavior of employees towards punctuality at workplace and can make appropriate decisions about the selection of employees at the time of making new intakes.

The main contribution of this paper is to analyze different parameters which can potentially contribute to absenteeism and then develop a model based on DNN to most accurately predict the absenteeism before employees are actually hired. To the best of our knowledge, none of the previous studies have considered such an extensive analysis of the absenteeism behaviour at workplace. This paper is the very first study of its kind and can be used as a baseline to further improve the prediction of absenteeism.

The rest of the paper is organized as follows. Different related studies are presented in Section 2. The proposed methodology is given in Section 3 along with a detailed analysis of the dataset used in the paper and different deep learning models that can predict the number of hours an employee is absent from work. Different optimization techniques that are used to improve the performance of machine learning models are also explored. Results collected from these learning algorithms are presented in Section 4, and the paper is concluded in Section 5.

2. Related Work

Absenteeism at workplace does have adverse impact on organizational environment and productivity. In recent years, researchers and scientists are getting more interest to

model these real-world problems using artificial intelligence-based techniques [8–11]. More recently in 2018, Gayathri conducted a research study for absenteeism at workplace [12] and proposed a model which can predict the number of hours an employee is absent from work. Four different categories of absenteeism are used. When the numbers of hours an employee's absenteeism are zero, then it is classified as a *Not ABSENT* class. When the number of absent hours is 1–16, it is classified as *DAYS*, and 17–56 absent hours are referred as *WEEK*, and above 56 absent hours as *MONTH*. Three different classification models (i.e., Naive Bayes, Decision Tree, and Multilayer Perceptron) are used in this research study. The highest accuracy is achieved by Multilayer Perceptron. However, there does not exist detailed analysis on the behavior of these classification models on the given dataset. It is also not clear how the Multilayer Perceptron model and the optimization techniques used are achieving highest accuracy. Furthermore, no further explanation of the model can be used in the industry to predict the behavior of employees and what action can be taken. A similar study was also conducted by Ferreira et al. in [13] to predict absenteeism at work. ANN is used in this research study to predict critical features as well as drop those features which are not contributing to enhancement of the accuracy of the model. Later, the model is trained with a reduced number of features. Similarly, another research study [14] also used ANN to predict absenteeism at work. However, detailed analysis of the classification model is not shown and deep learning algorithms are not explored. Application of a neuro fuzzy network in prediction of absenteeism at work is also applied in [15].

In another research study by Shandizi in [16], a pilot's absenteeism is predicted in an airline company. In the airline industry, crew costs are the second most important cost after fuel costs, and pilots are the most important airline crew. For airline companies, having a system that can predict pilot absenteeism can help to manage the operations. They are using the Decision Tree algorithm in building a decision support system to predict the number of hours a pilot will be absent and to make necessary arrangements to deal with the situation. This system is dealing only with pilot absenteeism and can be used in an airline industry.

In the literature, there are a number of research studies which are using machine learning and other data mining techniques to understand the hidden patterns in the data. These machine learning and data mining algorithms are extensively used as classification models to predict different patterns in dataset. For example, in [17], a Decision Tree is used as a data mining technique to predict the attendance pattern of employees. In this research study, a private company's data are used as a case study to test classification algorithms. Similarly, in [18–20], machine learning models are used to predict employee turnover. Also, in another study, data mining techniques are used to predict employee turnover [21]. In [22], machine learning algorithms (i.e., Random Forest) are used to predict dropout in high school students. In another study, the machine learning algorithm is used to identify students at risk of adverse academic

outcomes [23] and also, the data mining technique used to predict secondary school student performance is presented in [24].

Traditional machine learning algorithms known as Decision Tree, Gradient Boosted Tree, Random Forest, and Tree Ensemble are used for absenteeism in [25]. The paper reports an accuracy rate of 82% by the Gradient Boosted Tree, while Tree Ensemble performed the lowest in terms of the accuracy rate of 97%. Kang et al. in [26] have demonstrated the resource management and scheduling based on Stochastic-Petri Net Modeling and optimization for patients to make a sustainable healthcare system, which takes into account the absenteeism factor of medical staff. The study demonstrates that when the absenteeism factor of medical staff is taken into account, the performance of the healthcare system is improved significantly in terms of reduced waiting time for patients and improved operational sustainability. DNNs in the context of predicting image privacy have been studied in [27]. The study demonstrates that privacy of pictures uploaded by a user on social media is important, and hence, it is important that machine learning models can automatically predict whether the privacy of pictures uploaded on social media should be public or private. Deep learning algorithms along with PCA have been in making prediction of Stroke Patient Mortality in [28]. The paper demonstrates that the area under the curve of the proposed method based on deep learning was 83.48% and therefore can be effectively used by patients and doctors to prescreen for possible stroke. Education Data Mining (EDM) is a research field that focuses on the application of data mining, machine learning, and statistical methods to detect patterns in large collections of education data. Recently, Deep Learning is getting more attention in the field of EDM and has been thoroughly explored in [29]. In another study in [30], complex networks in stock market and stock price volatility pattern are combined with machine learning to predict stock price patterns. SVM and KNN algorithms are used and have achieved an accuracy of 70%.

All previous studies have applied machine learning and deep learning techniques to model diverse problems. However, to the best of our knowledge, there has been no effort to predict absenteeism behaviour of employees at early stage in an organization. Current state-of-the-art research papers are based on understanding the pattern of absenteeism at workplace and proposed different solutions to reduce the absenteeism rate. However, no research work is carried out in order to make prediction of absenteeism at early stage as proposed in this research paper. In addition to that, techniques of deep neural networks have not been explored on the problem of absenteeism. This research paper compares the performance of traditional machine learning and deep neural networks and concludes that deep neural network is a suitable model for prediction of absenteeism at early stage. Clearly, there is a research gap for modeling and predicting absenteeism behaviour of employees at workplace at the early stage of their hiring. The models presented in this

paper are developed as general models and can predict employee's behaviour at the time of hiring, whether the employee will be punctual or tend to be more frequently absent in future.

3. The Proposed Methodology

3.1. Data Analysis. In this research study, workplace absenteeism data are taken from the UCI Machine Learning repository (<https://archive.ics.uci.edu/ml/datasets.html>). The dataset contains a total of 20 different features and 740 samples. These data samples reflect the behaviour of employees towards punctuality at a courier company in Brazil. These features are *Reason for absence*, *Month of absence*, *Day of the week*, *Seasons*, *Transportation expense*, *Distance from residence to work*, *Service time*, *Age*, *Work load average/day*, *Hit target*, *Disciplinary failure*, *Education*, *Son*, *Social drinker*, *Social smoker*, *Pet*, *Weight*, *Height*, *Body mass index*, and *Absenteeism category*. We have designed the absenteeism category as *moderate* where the number of hours an employee is absent for 0–5 hours per month, and the other class is *excessive* when the number of hours an employee is absent for more than 5 hours. This extra number of hours of absence is calculated after the leaves that are allowed for employees as per organization's policy (i.e., one or two days per month and paternity/maternity leave). A relaxation of 5 hours is given in the model, only to differentiate between employees who happen to have a problem, compared to the employee who is habitually absent. But after 5 hours, organizations indicate their concerns about the number of hours employees are absent from work. The dataset is small, containing only 740 instances. It is not a problem for traditional machine learning algorithms such as SVM [31] or Logistic Regression [32], but Neural Network and particularly Deep Networks are data hungry. We demonstrate the results of Deep Networks using this small dataset and propose that such technique will work even better when there are millions of instances in the dataset.

In the dataset, some features have small values, e.g., in the range 0–10, and some features have large values as in the range of 100–1000. This can make the learning process slow. Therefore, data are standardized to enhance fast learning in the machine learning algorithms. After standardization, the values in the dataset will be in the range of $-1, \dots, 1$. The formula of standardization (also known as **z-score**) is expressed in equation (1), where X_i are all the samples for a given feature, \bar{X} is the average of all samples by the feature, and s is the standard deviation. After standardization, the values in the dataset will be in the range of -1 to 1 :

$$Z_i \leftarrow \frac{X_i - \bar{X}}{s}. \quad (1)$$

During the analysis of the dataset, it is observed that the *moderate* category has 67.23% data (i.e., 468/740) and the *excessive* category has 32.76% data (i.e., 272/740). This can lead the learning algorithms to be biased [33] towards the moderate category, as most of the time, the prediction will be moderate. In order to deal with the biasness, Synthetic Minority Oversampling Technique (SMOTE) presented by

Nitesh et al. in 2002 [34] and later made available as a tool to be used in Python in [35] is used. After SMOTE is applied to the dataset, the *moderate* category becomes 468/936 and the *excessive* category, 468/936. The sample size is increased from 740 to 936 as SMOTE uses oversampling.

The X dataset has a dimension of (20, 936), i.e., we have 20 dimensions in our dataset. It is not possible to visualize such as a high-dimensional data. But we can use dimensionality reduction techniques such as PCA [36] or t-SNE [37, 38] to reduce high-dimensional data to two or three dimensions and then visualize it. Figure 1 shows the visualization of the dataset in 3D and gives an intuition of the learning process in the machine learning algorithm.

3.2. Learning Models. In this paper, the dataset is explored on (Shallow) Neural Network and Deep Neural Networks. A simple NN can be considered as an extended version of Logistic Regression [32] that is used to classify classes in a dataset. In a Shallow Neural Network [39], there is an input layer that contains all the features of all training examples. In the hidden layer, training data are multiplied with weights and an activation function is applied. Generally, at the hidden layer, tanh, ReLU, leaky ReLU, Randomized ReLU, or other activation functions are used. At the output layer, sigmoid is used in binary classification problems and softmax is used in multiclass classification problems. In the experiments of this paper, sigmoid activation function is used as the classification problem is a binary classification problem. There are different types of cost functions such as MSE or RMSE. However, these types of cost functions when used in classification problems result in a nonconvex function of all loss functions and will have multiple local minima. Therefore, cross-entropy loss function is used as the resultant sum of all cost functions is in a convex shape and therefore does not have the problem of local minima. A block diagram of a Shallow Neural Network that is used for the classification of the dataset is shown in Figure 2.

We will explain the process of learning for absenteeism prediction. There are 20 features, where 19 are input features and 1 is output feature, i.e., whether a person will fall in the *moderate* or *excessive* category. In order to perform training, we store all data in a matrix. We have 936 instances of absenteeism, and therefore, the size of the input matrix represented by X is 936×19 . In order to train NN, we need to provide a matrix of the weights with the same size as input features. In our case, there are 10 units in the first layer and the size of the weight matrix is 19×10 . We initialize these weights randomly using *GlorotUni forminitializer*. We also need to provide a bias represented by b . The formula of this multiplication is shown in equation (2), where W_1 shows the weights for the hidden layer, b_1 shows the bias, and X represents the input matrix. We also have to perform a nonlinear function ReLU [?], which is computed as $ReLU(z) = \max(0, z)$:

$$\begin{aligned} Z_1 &\leftarrow W_1^T \times X + b_1, \\ A_1 &\leftarrow ReLU(Z_1). \end{aligned} \quad (2)$$

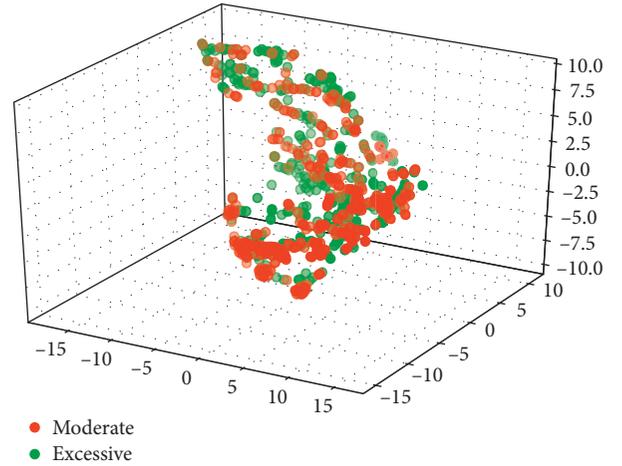


FIGURE 1: Visualization of the dataset in 3D.

For the output layer, we multiply the output of the hidden layer with different weights. Let us say we have 10 units in the hidden layer and one unit in the output layer; then, the dimension of the weight matrix is 10×1 . We also need to add a bias at this layer. The calculation performed at the output layer is shown in equation (3), where W_2 and b_2 show the weight and bias for the output layer and A_1 is the input vector. At the output layer, sigmoid [40] is computed as $\text{sigmoid}(Z) = 1/(1 + e^z)$:

$$Z_2 \leftarrow W_2^T \times A_1 + b_2, \quad (3)$$

$$A_2 \leftarrow \text{sigmoid}(Z_2),$$

$$\mathcal{L}(A_2, Y) \leftarrow -\frac{1}{m} \sum_{i=0}^m Y_i \log(A_2). \quad (4)$$

During the training phase of the NN, the prediction is made, represented by A_2 , as shown in equation (3). Then, the loss is computed comparing the predicted values with actual values. We are using *binarycrossentropy* loss as shown in equation (3), where m represents the number of samples and Y shows the actual output values. During the backpropagation process, the derivative of the loss is taken for output layer and hidden layer, and weights are updated using optimization techniques, such as Gradient Descent [41], Gradient Descent with momentum [42], RMSProp [43], and Adam [44]. The backpropagation with Gradient Descent is shown in equation (5). The algorithm of training of the NN model for absenteeism prediction is given in Algorithm 1, where α is the learning rate:

If the number of layers is increased, then the learning process is called deep learning and the network is generally called Deep Neural Network. A general consensus [45] is that when there are more than two hidden layers, then the era of deep learning begins. The whole process of training a deep learning framework can be explained in Figure 3. The forward propagation and backward propagation are performed for a number of iterations, until the cost cannot be further decreased. The pseudocode of the algorithm of deep learning for absenteeism prediction is

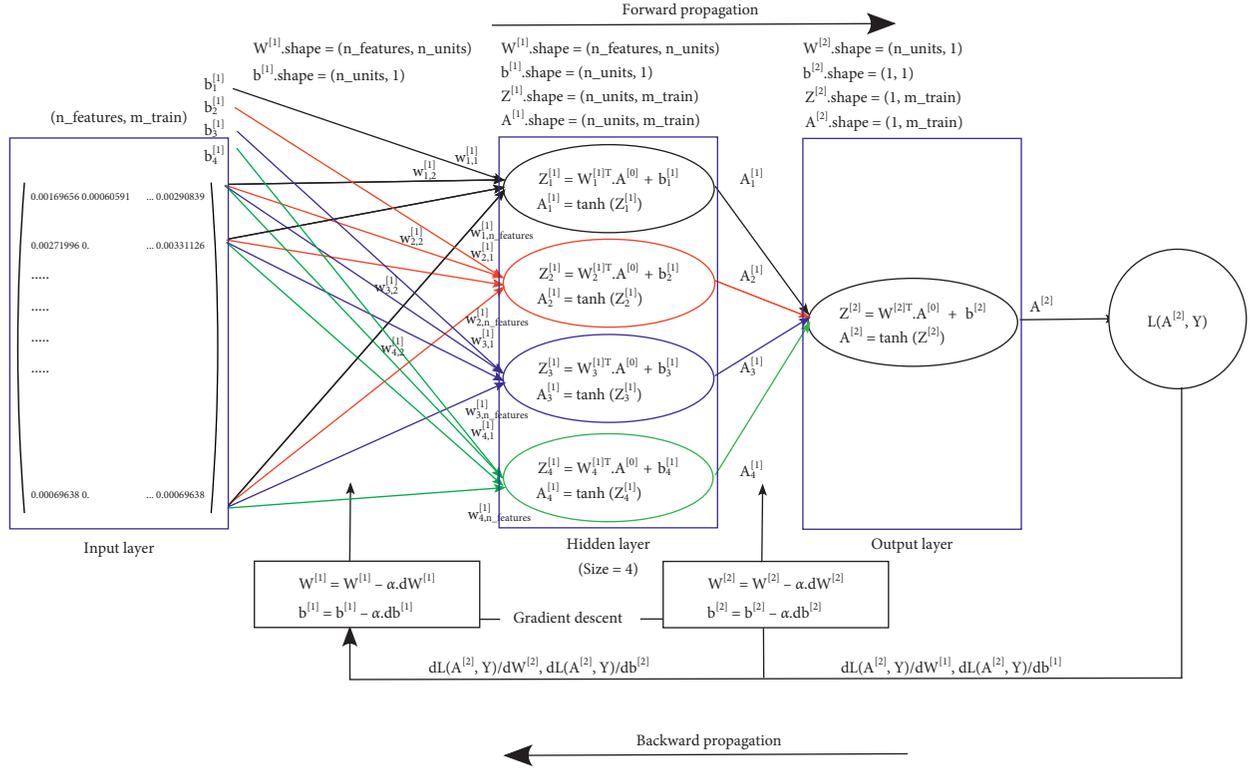


FIGURE 2: Shallow Neural Network is created when Logistic Regression is applied multiple times. We also show the forward propagation to compute loss and backward propagation to update hyperparameters $W^{[1]}$, $W^{[2]}$, $b^{[1]}$, and $b^{[2]}$. The forward propagation and backward propagation are performed for a number of iterations, until we reach to the global minima of the sum of cross-entropy loss for all features.

```

 $W_1, b_2, W_1, b_2 \leftarrow$  Glorot Uniform initializer
while  $i \leq \text{num\_iteration}$  do
 $Z_1 \leftarrow W_1^T \times X + b_1$ 
 $A_1 \leftarrow \text{ReLU}(Z_1)$  where  $\text{ReLU}(Z) = \max(0, z)$ 
 $Z_2 \leftarrow W_2^T \times A_1 + b_2$ 
 $A_2 \leftarrow \text{sigmoid}(Z_2)$  where  $\text{sigmoid}(z) = 1 / (1 + e^z)$ 
 $\mathcal{L}(A_2, Y) \leftarrow -1/m \sum_{i=0}^m Y_i \log(A_2)$ 
 $W_1 \leftarrow W_1 - \alpha \partial \mathcal{L} / \partial W_1$ 
 $b_1 \leftarrow b_1 - \alpha (\partial \mathcal{L} / \partial b_1)$ 
 $W_2 \leftarrow W_2 - \alpha (\partial \mathcal{L} / \partial W_2)$ 
 $b_2 \leftarrow b_2 - \alpha (\partial \mathcal{L} / \partial b_2)$ 
end while

```

ALGORITHM 1: Training in Shallow Neural Network.

given in Algorithm 2. After training, the learning algorithm is used to make prediction on the Test dataset. The values of \hat{Y} or A are probabilities in the range 0-1. In the classification of the dataset, the output class is either moderate or excessive. Therefore, the result of A is adjusted such that if the probability is below 0.5, then *moderate* is produced; otherwise, *excessive* is produced, as shown in

$$y_{\text{prediction}}^{(i)} \leftarrow \begin{cases} \text{excessive,} & \text{if } A^{(i)} > 0.5, \\ \text{moderate,} & \text{otherwise.} \end{cases} \quad (5)$$

$$\begin{aligned} W_1 &\leftarrow W_1 - \alpha \frac{\partial \mathcal{L}}{\partial W_1}, \\ b_1 &\leftarrow b_1 - \alpha \frac{\partial \mathcal{L}}{\partial b_1}, \\ W_2 &\leftarrow W_2 - \alpha \frac{\partial \mathcal{L}}{\partial W_2}, \\ b_2 &\leftarrow b_2 - \alpha \frac{\partial \mathcal{L}}{\partial b_2}. \end{aligned} \quad (6)$$

Various optimization techniques can be applied in Neural Network models to further improve the performance of the

```

 $W^{[1..L]} \leftarrow$  Glorot Uniform initialize
 $b^{[1..L]} \leftarrow$  Glorot Uniform initialize
while  $i \leq \text{num\_iteration}$  do
   $k \leftarrow 1$ 
  while  $j \leq L$  do
     $Z^{[j]} \leftarrow W^{[j]T} \cdot A^{[j-1]} + b^{[j]}$ 
     $A^{[j]} \leftarrow g(Z^{[j]})$  where  $g(Z) = \max(0, z)$ 
    increment  $j$  by 1
  end while
   $\mathcal{L}(A^{[L]}, Y) \leftarrow -(1/m) \sum_{i=0}^m Y_i \log(A_i^{[L]})$ 
   $k \leftarrow L$ 
  while  $k \geq 0$  do
     $W^{[k]} \leftarrow W^{[k]} - \alpha (\partial \mathcal{L} / \partial W^{[k]})$ 
     $b^{[k]} \leftarrow b^{[k]} - \alpha (\partial \mathcal{L} / \partial b^{[k]})$ 
    decrement  $k$  by 1
  end while
end while

```

ALGORITHM 2: Training in Deep Neural Network.

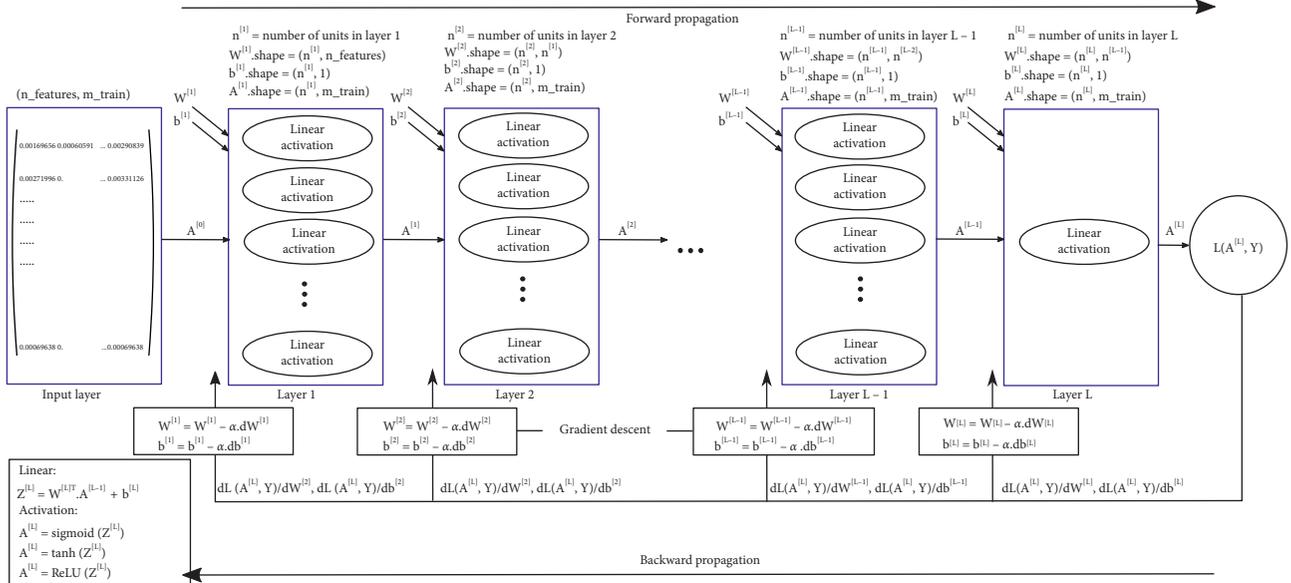


FIGURE 3: In Deep Neural Network, there are multiple hidden layers instead of a single layer of a Shallow Neural Network. In forward propagation, we compute linear and activation for all hidden layers and then compute the loss at the output layer. Then, through backpropagation, the derivative of loss with respect to weight and bias terms is calculated for all the layers in order to have optimized values of weight and bias terms. In the learning process, the forward propagation and backward propagation are performed for a number of iterations, until there is no further reduction in cost. $W^{[L]}$ and $b^{[L]}$ represent the weight and bias for layer L . $A^{[L-1]}$ is the input parameter to Layer L , and $A^{[L]}$ is the output produced from layer L . $n^{[L]}$ represents the number of units (also known as Neurons) in a given layer. α is the learning rate for the gradient descent algorithm.

algorithm. In Neural Networks and Deep Neural Networks, zero initialization does not perform symmetry breaking [46]. It has been experimentally observed that random initialization breaks the symmetry and gives better accuracy. There are different initialization methods such as Xavier initializer (also known as Glorot Uniform Initializer) [47] or He initializer [48] that perform even better than random initialization. Gradient Descent [49] can be slow to reach the global minimum of the sum of loss function. There are better optimization algorithms that can converge quickly to the global minimum. These

algorithms are Stochastic Gradient Descent [50], Mini-batch Gradient Descent [51], Gradient Descent with momentum [42], RMSProp [52], and Adam Optimization [53, 54].

In this paper, we are using Adam Optimization for the learning process as it is one of the most effective optimization algorithms for training in Deep Neural Networks. Adam Optimization can be expressed mathematically in equation (7). Here, $v_{dW^{[l]}}^{\text{corrected}}$ stores the exponentially weighted average of past gradients with bias correction for layer l , $s_{dW^{[l]}}^{\text{corrected}}$ calculates exponentially weighted average of

the squares of the past gradients for layer l , (β_1) and (β_2) are hyperparameters that control the two exponentially weighted averages, α is the learning rate, t counts the number of steps taken of Adam optimization, l means the number of layers, and ϵ is a tiny value to avoid divide by zero error:

$$\begin{aligned} v_{dW^{[l]}} &\leftarrow \beta_1 v_{dW^{[l]}} + (1 - \beta_1) \frac{\partial \mathcal{L}}{\partial W^{[l]}}, \\ v_{dW^{[l]}}^{\text{corrected}} &\leftarrow \frac{v_{dW^{[l]}}}{1 - (\beta_1)^t}, \\ s_{dW^{[l]}} &\leftarrow \beta_2 s_{dW^{[l]}} + (1 - \beta_2) \left(\frac{\partial \mathcal{L}}{\partial W^{[l]}} \right)^2, \\ s_{dW^{[l]}}^{\text{corrected}} &\leftarrow \frac{s_{dW^{[l]}}}{1 - (\beta_2)^t}, \\ W^{[l]} &\leftarrow W^{[l]} - \alpha \frac{v_{dW^{[l]}}^{\text{corrected}}}{\sqrt{s_{dW^{[l]}}^{\text{corrected}} + \epsilon}}. \end{aligned} \quad (7)$$

Overfitting is caused when the model is trained well during training but does not generalize well during testing. A standard way to avoid overfitting is called regularization [55] or Dropout [56]. Learning rate is another hyperparameter that can be optimized. In this paper, the learning rate decay algorithm [57] is used that starts from one learning rate and as the algorithm is converging, the learning rate is decreased.

The main objective of this research work is the novel technique of identifying different parameters that can contribute to the absenteeism, preprocess the data to be processed by deep learning algorithms efficiently, and then devise a deep learning algorithm with most recent optimization techniques that can make prediction of absenteeism with reasonable accuracy. Even though many researchers have worked on absenteeism and have demonstrated to find Artificial Intelligence-based solutions for it, no one has studied an effective mechanism of understanding factors of absenteeism using deep learning, which is becoming very popular recently with the increased data and increased computational [58–62] power. According to the knowledge of the authors, no comprehensive work is dedicated to absenteeism prediction using deep learning algorithms. Therefore, it is sensible to study the problem of absenteeism from the perspective of deep learning to demonstrate the full potential of Deep Neural Network.

4. Results

In Shallow Neural Networks, there are 100 units in the hidden layer with an initial learning rate of 0.01 with adaptive learning strategy β_1 as 0.9 and β_2 as 0.999, 1000 number of epochs, ReLU activation function in the hidden layer and sigmoid activation function at the output layer, and 10-fold cross validation. In the Deep Neural Network, all these parameters remain the same, except that there are six hidden layers instead of a single hidden layer. There are 200 units in the first layer, 150 units in the second layer, 100 in the third layer, 50 in the fourth layer, 10 in the fifth layer, and 5 in the sixth layer. These hyperparameters are selected based on tuning of the training and validation datasets. The most

commonly used technique for choosing hyperparameters are grid search and random search. It has been experimentally proved that random search saves much more time in selecting optimized hyperparameters. In this paper, random-search technique is used to search for optimized hyperparameters.

4.1. Accuracy. The machine learning models are trained on Train data, and then, Dev dataset is used to tune hyperparameters and retrain. Then, the trained model is used to make predictions on the Test dataset. The accuracy of Dev and Test dataset for Shallow Neural Network and Deep Neural Network is shown in Figure 4. The accuracy achieved by Shallow neural network is 79.8%. As the number of layers is increased, make it a deep neural network; then, the accuracy is further increased up to 97.5%. Although different techniques such as regularization and Dropout are used, for deep learning to be effective, the role of large data containing millions of instances cannot be ignored.

4.2. Precision, Recall, and F1-Score. The comparison of Precision, Recall, and F1-score in the classification models is shown in Figure 5. Deep Neural Networks has the highest precision, recall, and F1-score compared to a Shallow Neural Network and is therefore the best model to be used for this dataset.

4.3. Recursive Operating Characteristic (ROC) Curve. The performance of a model can also be determined from ROC by analysing the Area under the Curve (AUC). A better model will pass through the upper left corner (100% sensitivity, 100% specificity) and will have a higher overall accuracy of the test [63]. The ROC curve for the classification models presented in this paper on the test dataset is shown in Figure 6. The ROC of Shallow Neural Network is shown in Figure 6(a). The ROC achieved by Deep Neural Network is shown in Figure 6(b) and covers more area than the single-layer NN, demonstrating that DNN is the most suitable model for absenteeism prediction at workplace.

4.4. Confusion Matrix. The performance of a model can also be computed by analysing the confusion matrix. The confusion matrix computed for NN and DNN on the test dataset is shown in Figure 7. The more the values on the diagonal, the higher the accuracy of the model. The confusion matrix for NN is given in Figure 7(a) and shows that there are lower values as compared to the confusion matrix for DNN given in Figure 7(b). This experiment further demonstrates that DNN is the most suitable learning algorithm for predicting absenteeism at workplace.

4.5. Contribution of Individual Features. The model uses 19 input features to predict the absenteeism category as *moderate* or *excessive*. In Neural Networks, these input features are assigned random weights at the first iteration, and then with each iteration, these weights are updated through backpropagation with an Adam optimizer. When the model is trained, the optimized values for all weights are

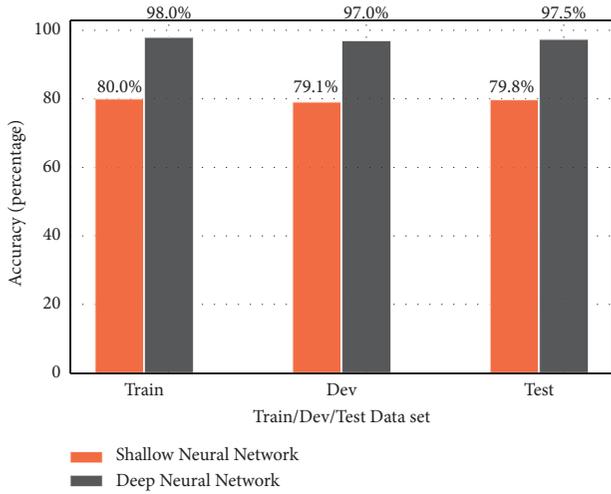


FIGURE 4: Performance/accuracy of learning models in the dataset.

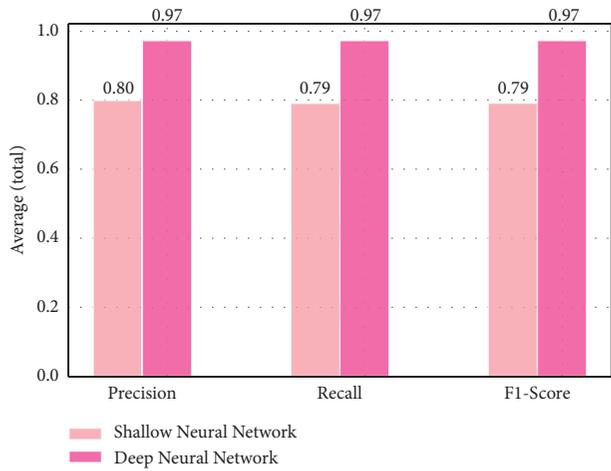


FIGURE 5: Comparison of Precision, Recall, and F1-Score in Neural Network models.

reached and then these weights are used to make prediction on Dev/Test dataset. In the Deep Neural Network, there are 200 units in the first hidden layer, and therefore, the shape of weights at the input layer is (19, 200), i.e., there are 200 weights for each feature. The average of these values is taken, and hence, 19 weights are obtained. Then, these weights are converted to percentage in order to know their contribution in the prediction of output. Some weights have negative values, and their contribution is made as nonnegative. These contribution percentages by each individual feature are shown in Figure 8. The features that can contribute more to the absenteeism by employee are *Reasons of absence*, *Seasons*, *Transportation expense*, *Distance from residence to work*, *Disciplinary failure*, *Social smoker*, and *Body mass index*.

4.6. Comparison with Other Approaches. The proposed approaches based on Deep Neural Network have been compared with other machine learning algorithms, i.e., Decision Tree, support vector machine (SVM), and Random Forest. The performance of these algorithms in terms of accuracy,

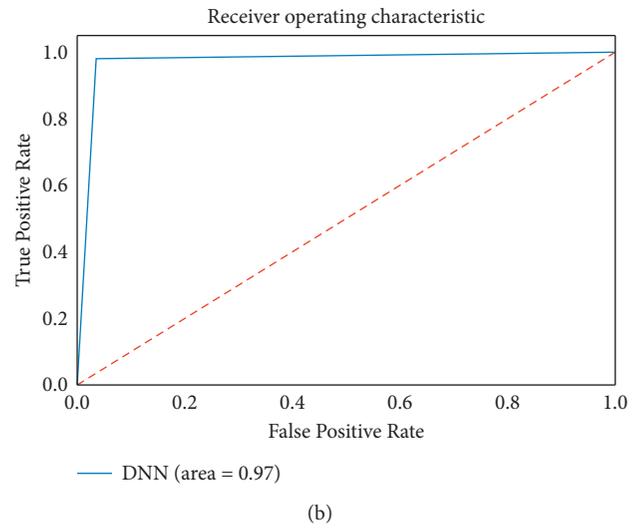
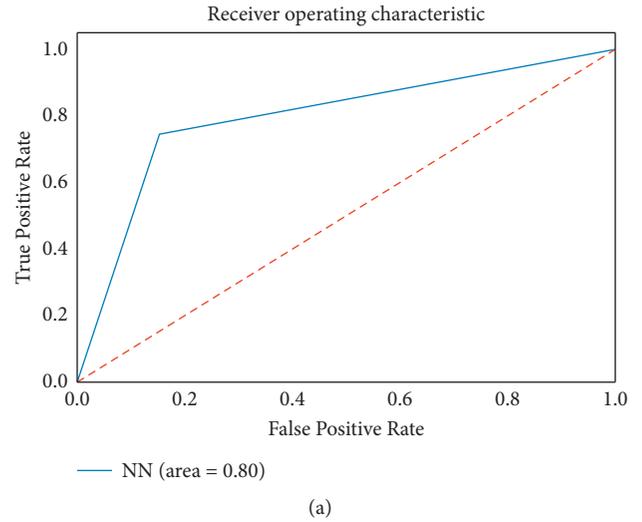


FIGURE 6: ROC of the test dataset computed in (a) Shallow Neural Network and (b) Deep Neural Network.

True Positive Rate, False Positive Rate, F1-Score, and ROC curve area is shown in Table 1.

In the dataset, 28 different reasons of absences are given, such as absence because of sickness and absence because of emergency. The results achieved by Decision Tree J48 that takes different reasons of absences and compute the probability of person in the moderate or excessive category are shown in Table 2.

The proposed method based on Deep Neural Network achieves an accuracy of 97.5%, whereas Decision Tree achieves an accuracy of 82.83%, SVM achieves an accuracy of 84.32%, and Random Forest achieves an accuracy of 82.43%. This comparison demonstrates the usefulness of the proposed architecture. Although Decision Tree archives an accuracy of 82%, it identifies attributes such as *Reason for absence*, *Hit target*, *Distance from residence to work*, *Social drinker*, *Age*, and *Height* as important attributes that contribute significantly towards absenteeism at workplace. Similar attributes are identified as important attributes by the proposed model as shown in Figure 8. The comparison of the proposed model

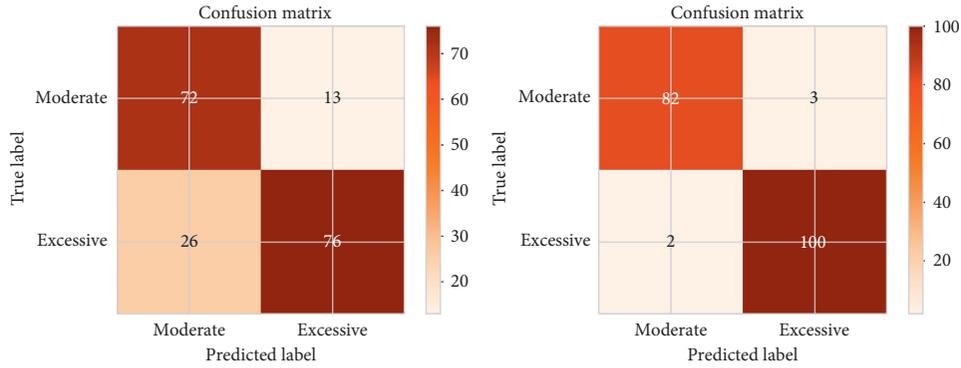


FIGURE 7: Confusion Matrix of the test dataset computed by (a) Shallow Neural Network and (b) Deep Neural Network.

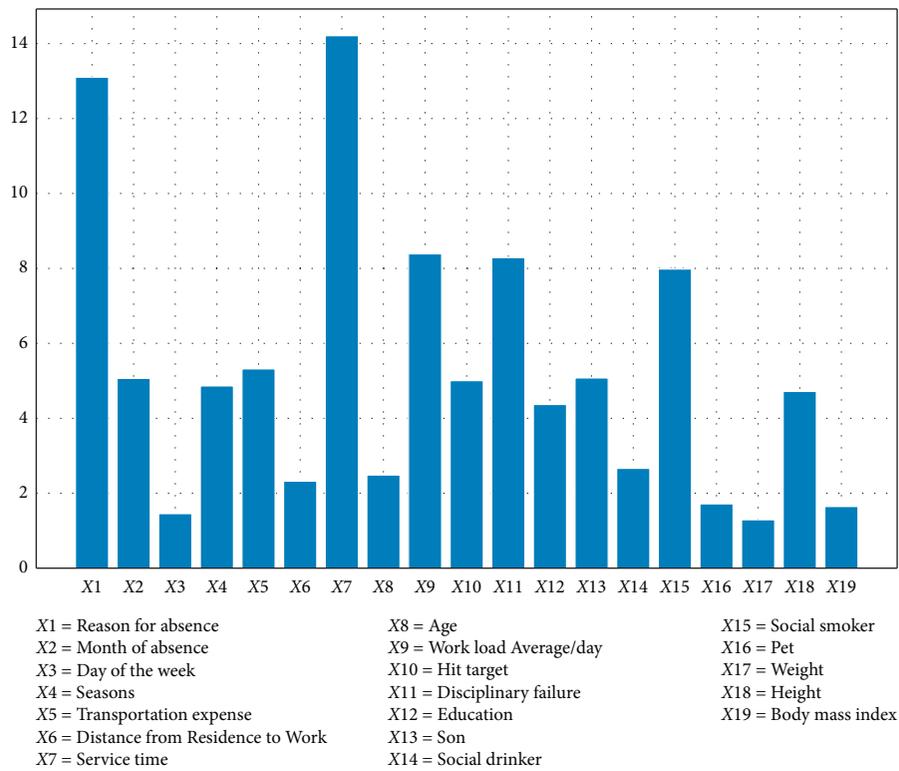


FIGURE 8: Contribution of individual features in the Deep Neural Network models.

TABLE 1: Accuracy comparison in different machine learning models.

Method	Accuracy (%)	TP rate	FP rate	F1-score	ROC area
Decision Tree	82.83	0.828	0.204	0.828	0.851
SVM	84.32	0.843	0.154	0.845	0.845
Random Forest	82.43	0.824	0.228	0.822	0.886

with traditional machine learning algorithms validates the finding in this paper. The proposed approach finds *Body mass* attribute as an important factor in absenteeism, and the Decision Tree finds *Age* and *Height* as important factors which are related to *Body mass*. This comparison with other machine learning approaches demonstrates and validates the effectiveness of the proposed approach.

4.7. Applicability of the Proposed Models in Real World. The learning models presented in this paper are used to predict the behavior of employees towards punctuality at workplace during an early stage of conducting interviews for hiring. The results of the models which are presented in Section 4 can be implemented in the real world. For instance, if an organization intends to deploy them, they

TABLE 2: Reasons of absence that results into moderate or excessive categories using J48 Decision Tree.

Reason for absence = 0: moderate (43.0)
Reason for absence = 1: excessive (16.0/2.0)
Reason for absence = 2: excessive (1.0)
Reason for absence = 3: excessive (1.0)
Reason for absence = 4: moderate (2.0/1.0)
Reason for absence = 5: excessive (3.0/1.0)
Reason for absence = 6: excessive (8.0/1.0)
Reason for absence = 7: excessive (15.0/7.0)
Reason for absence = 8
> Hit target ≤ 95 : excessive (3.0)
> Hit target > 95 : moderate (3.0)
Reason for absence = 9: excessive (4.0)
Reason for absence = 10: excessive (25.0/4.0)
Reason for absence = 11
> Month of absence ≤ 3 : excessive (7.0)
> Month of absence > 3 : moderate (19.0/7.0)
Reason for absence = 12
> Distance from residence to work ≤ 27 : excessive (5.0/1.0)
> Distance from residence to work > 27 : moderate (3.0)
Reason for absence = 13
> Social drinker = 0: moderate (21.0/7.0)
> Social drinker = 1: excessive (34.0/3.0)
Reason for absence = 14
> Social drinker = 0: moderate (6.0)
> Social drinker = 1
> Age ≤ 41 : excessive (10.0/1.0)
> Age > 41 : moderate (3.0)
Reason for absence = 15: excessive (2.0)
Reason for absence = 16: moderate (3.0)
Reason for absence = 17: excessive (1.0)
Reason for absence = 18: excessive (21.0/4.0)
Reason for absence = 19
> Height ≤ 175 : excessive (30.0/1.0)
> Height > 175 : moderate (10.0/4.0)
Reason for absence = 20: moderate (0.0)
Reason for absence = 21: excessive (6.0/2.0)
Reason for absence = 22: excessive (38.0/4.0)
Reason for absence = 23: moderate (149.0/13.0)
Reason for absence = 24: excessive (3.0)
Reason for absence = 25: moderate (31.0/5.0)
Reason for absence = 26: excessive (33.0/8.0)
Reason for absence = 27: moderate (69.0)
Reason for absence = 28: moderate (112.0/8.0)

can use either the existing dataset or even add their own data to train the proposed models. When they are hiring new employees, they can collect the required data from the model to predict the behavior of candidates towards punctuality. The algorithm will predict candidates in the *moderate* category or *excessive* category. The employer can then make a decision to exclude all candidates in the *excessive* category from interviews and forward only those candidates in the *moderate* category to the next evaluation process in hiring. This way the organization gets benefits from using Deep Neural Network to filter out those candidates who will not be punctual at work at the early stage of hiring. The process of how the organization can use the DNN-based model in the selection of new candidates is shown in Figure 9.

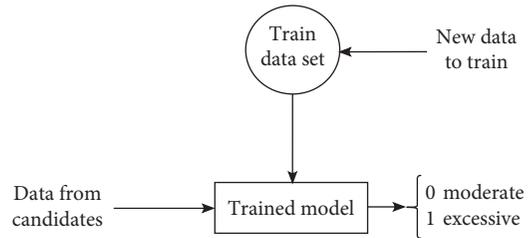


FIGURE 9: Usage of the DNN-based model for absenteeism prediction in real world.

5. Conclusion

Organizations are very concerned about the behavior of employees towards punctuality at work, as organizations can make progress only if employees are putting enough number of hours to work. But those employees who are habitually late and looking for excuses to steal time from work can be a real trouble for organizations. In order to avoid the critical issue of absenteeism where organizations have to confront those who are stealing time from work, a structured methodology using machine learning and deep neural network models have been presented which can determine the behaviour of such employees towards punctuality at work at the very early stage of their hiring. The results obtained have demonstrated higher accuracy by the proposed model and show a great potential to be scaled by using a big dataset for real-world problems. The dataset used in this research study belongs to a courier company in Brazil, and the dataset contains features which reflect human behaviours; such behaviours may correlate differently in other geographic locations. The proposed model can be extended to other global locations by adapting to local employee's behavioural features. Then; the cultural differences and demographic issues will not affect the model efficacy. The proposed models can be trained with the local dataset first. Afterwards, the model can learn and produce more reliable and accurate results. The most important contribution of the paper is to analyze the parameters used for employee selection and devise a deep learning-based model to make prediction of absenteeism behavior in employees. It also highlights the contributions of different factors that can possibly result in absenteeism of employees and is a good source for organizations to further look into those areas and take necessary actions to promote a healthy environment at workplace.

Data Availability

The data used to support the findings of this study have been uploaded to the GitHub repository (<https://github.com/mirfanud/Absenteeism.git>).

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through the research group under no. RG- 1438-089.

References

- [1] M. C. Sturman, *Multiple Approaches to Absenteeism Analysis*, Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies, London, UK, 1996.
- [2] F. Cucchiella, M. Gastaldi, and L. Ranieri, "Managing absenteeism in the workplace: the case of an Italian multiutility company," *Procedia-Social and Behavioral Sciences*, vol. 150, pp. 1157–1166, 2014.
- [3] C. A. M. Roelen, P. C. Koopmans, R. Hoedeman, U. Bultmann, J. W. Groothoff, and J. J. L. van der Klink, "Trends in the incidence of sickness absence due to common mental disorders between 2001 and 2007 in The Netherlands," *The European Journal of Public Health*, vol. 19, no. 6, pp. 625–630, 2009.
- [4] S. D. Russo, M. Miraglia, L. Borgogni, and G. Johns, "How time and perceptions of social context shape employee absenteeism trajectories," *Journal of Vocational Behavior*, vol. 83, no. 2, pp. 209–217, 2013.
- [5] M. Tkáč and R. Verner, "Artificial neural networks in business: two decades of research," *Applied Soft Computing*, vol. 38, pp. 788–804, 2016.
- [6] A. Novikov, D. Podoprakhin, A. Osokin, and D. P. Vetrov, "Tensorizing neural networks," in *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds., pp. 442–450, Curran Associates, Inc., New York, NY, USA, 2015.
- [7] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," in *Proceedings of the NIPS Deep Learning and Representation Learning Workshop*, New York, NY, USA, 2015.
- [8] D. K. Pradhan, J. Chakraborty, and S. Nandi, "Applications of machine learning in analysis of citation network," in *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, CoDS-COMAD '19*, pp. 330–333, New York, NY, USA, 2019.
- [9] G. Bonaccorso, *Machine Learning Algorithms: A Reference Guide to Popular Algorithms for Data Science and Machine Learning*, Packt Publishing, Birmingham, UK, 2017.
- [10] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, Pearson Education, London, UK, 2nd edition, 2003.
- [11] V. S. Kodogiannis, T. Pachidis, and E. Kontogianni, "An intelligent based decision support system for the detection of meat spoilage," *Engineering Applications of Artificial Intelligence*, vol. 34, pp. 23–36, 2014.
- [12] T. Gayathri, "Data mining of absentee data to increase productivity," *International Journal of Engineering and Techniques*, vol. 4, no. 3, pp. 478–180, 2018.
- [13] R. P. Ferreira, A. Martiniano, D. Napolitano, E. B. P. Farias, and R. J. Sassi, "Artificial neural network and their application in the prediction of absenteeism at work," *International Journal of Recent Scientific Research*, vol. 9, no. 1, pp. 2332–2334.
- [14] H. Trivedi, *Explaining Absenteeism at Workplace Predicted by a Neural Network*, Springer, Berlin, Germany, 2010.
- [15] A. Martiniano, R. P. Ferreira, R. J. Sassi, and C. Affonso, "Application of a neuro fuzzy network in prediction of absenteeism at work," in *Proceedings of the 7th Iberian Conference on Information Systems and Technologies (CISTI 2012)*, vol. 1–4, Madrid, Spain, June 2012.
- [16] A. H. H. Shandizi, *Prediction of Pilot's Absenteeism in an Airline Company*, Universite De Montreal, Montreal, Canada, 2014.
- [17] N. N. Qomariyah and Y. G. Sucahyo, "Employees' attendance patterns prediction using classification algorithm case study: a private company in Indonesia," *Int'l Journal of Computing, Communications & Instrumentation Engg*, vol. 1, no. 1, pp. 2349–1477, 2014.
- [18] R. Punnoose and P. Ajit, "Prediction of employee turnover in organizations using machine learning algorithms," *International Journal of Advanced Research in Artificial Intelligence*, vol. 5, no. 9, 2016.
- [19] E. Ribes, K. Touahri, and B. Perthame, "Employee turnover prediction and retention policies design: a case study," *CoRR*, vol. 10, 2017.
- [20] P. Rohit and P. Ajit, "Prediction of employee turnover in organizations using machine learning algorithms," *International Journal of Advanced Research in Artificial Intelligence*, vol. 5, p. 10, 2016.
- [21] A. M. Esmaeeli Sikaroudi, R. Ghousi, and A. Sikaroudi, "A data mining approach to employee turnover prediction (case study: arak automotive parts manufacturing)," *Journal of Industrial and Systems Engineering*, vol. 8, no. 4, pp. 106–121, 2015.
- [22] N.-B. Āzara, R. Halland, C. Igel, and S. Alstrup, "High-school dropout prediction using machine learning: a Danish large-scale study," in *Proceedings of the ESANN 2015*, M. Verleysen, Ed., pp. 319–324, i6doc.com, Bruges, Belgium, 2015.
- [23] H. Lakkaraju, E. Aguiar, C. Shan et al., "A machine learning framework to identify students at risk of adverse academic outcomes," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15*, ACM, New York, NY, USA, 2015.
- [24] P. Cortez and A. Silva, *Using Data Mining to Predict Secondary School Student Performance*, EUROSIS, London, UK, 2008.
- [25] Z. Wahid, A. K. M. Z. Satter, A. Al Imran, and T. Bhuiyan, "Predicting absenteeism at work using tree-based learners," in *Proceedings of the 3rd International Conference on Machine Learning and Soft Computing, ICMLSC 2019*, ACM, New York, NY, USA, 2019.
- [26] C. W. Kang, M. Imran, M. Omair, W. Ahmed, M. Ullah, and B. Sarkar, "Stochastic-petri net modeling and optimization for outdoor patients in building sustainable healthcare system considering staff absenteeism," *Mathematics*, vol. 7, no. 6, p. 499, Jun 2019.
- [27] A. Tonge and C. Caragea, "Image privacy prediction using deep features," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI'16*, pp. 4266–4267, AAAI Press, New York, NY, USA, 2016.
- [28] S. Cheon, J. Kim, and J. Lim, "The use of deep learning to predict stroke patient mortality," *International Journal of Environmental Research and Public Health*, vol. 16, no. 11, p. 1876, 2019.
- [29] A. Hernández-Blanco, B. Herrera-Flores, D. Tomás, and B. Navarro-Colorado, "A systematic review of deep learning approaches to educational data mining," *Complexity*, vol. 298, 2019.

- [30] H. Cao, T. Lin, Y. Li, and H. Zhang, "Stock price pattern prediction based on complex network and machine learning," *Complexity*, vol. 2019, 2019.
- [31] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and Their Applications*, vol. 13, no. 4, pp. 18–28, 1998.
- [32] M. Collins, R. E. Schapire, and Y. Singer, "Logistic regression, adaboost and bregman distances," *Machine Learning*, vol. 48, pp. 253–285, 2002.
- [33] A. B. Owen, "Infinitely imbalanced logistic regression," *Journal of Machine Learning Research*, vol. 8, pp. 761–773, 2007.
- [34] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [35] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, no. 17, 2017.
- [36] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.
- [37] L. van der Maaten and G. Hinton, *Visualizing Data Using T-Sne*, Prentice Hall PTR, Saddle River, NJ, USA, 2008.
- [38] L. Van Der Maaten, "Accelerating t-sne using tree-based algorithms," *Journal of Machine Learning Research*, vol. 15, pp. 3221–3245, 2014.
- [39] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 2nd edition, 1998.
- [40] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals, and Systems*, vol. 2, no. 4, pp. 303–314, 1989.
- [41] S. Ruder, "An overview of gradient descent optimization algorithms," 2016.
- [42] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Networks*, vol. 12, no. 1, pp. 145–151, 1999.
- [43] T. Tieleman and G. Hinton, "Lecture 6.5—RmsProp: divide the gradient by a running average of its recent magnitude," *COURSERA: Neural Networks for Machine Learning*, vol. 245, 2012.
- [44] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *Proceedings of the 3rd International Conference for Learning Representations*, San Diego, CA, USA, 2015.
- [45] Y. B. Goodfellow and A. Courville, *Deep Learning*, The MIT Press, Cambridge, MA, USA, 2016.
- [46] S. Masood and P. Chandra, "Training neural network with zero weight initialization," in *Proceedings of the CUBE International Information Technology Conference, CUBE '12*, pp. 235–239, ACM, New York, NY, USA, 2012.
- [47] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the JMLR W&CP Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2010)*, vol. 9, pp. 249–256, New York, NY, USA, May 2010.
- [48] X. Z. He and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), ICCV '15*, pp. 1026–1034, IEEE Computer Society, Washington, DC, USA, 2015.
- [49] J. B. Mason and M. Frean, "Boosting algorithms as gradient descent," *Advances in Neural Information Processing Systems*, vol. 12, pp. 512–518, 2000.
- [50] M. W. Zinkevich, L. Li, and A. J. Smola, "Parallelized stochastic gradient descent," in *Advances in Neural Information Processing Systems 23*, J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, Eds., pp. 2595–2603, Curran Associates, Inc., New York, NY, USA, 2010.
- [51] T. Li, Y. Chen and A. J. Smola, "Efficient mini-batch training for stochastic optimization," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14*, pp. 661–670, ACM, New York, NY, USA, 2014.
- [52] T. Tieleman and G. Hinton, "RMSprop gradient optimization."
- [53] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *CoRR*, vol. 1412, p. 6980, 2014.
- [54] T. Chilimbi, Y. Suzue, J. Apacible, and K. Kalyanaraman, "Project adam: building an efficient and scalable deep learning training system," in *Proceedings of the 11th USENIX Conference on Operating Systems Design and Implementation, OSDI'14*, pp. 571–582, USENIX Association, Berkeley, CA, USA, 2014.
- [55] S. M. Kakade, S. Shalev-Shwartz, and A. Tewari, "Regularization techniques for learning with matrices," *Journal of Machine Learning Research*, vol. 13, pp. 1865–1890, 2012.
- [56] G. H. Srivastava, A. I. Krizhevsky, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, 2014.
- [57] T. Schaul, S. Zhang, and Y. LeCun, "No more pesky learning rates," in *Proceedings of the 30th International Conference on Machine Learning*, vol. 28, S. Dasgupta and D. McAllester, Eds., pp. 343–351, Atlanta, Georgia, USA, 2013.
- [58] M. Fatima, A. Baig, and I. Uddin, "Reliable and energy efficient mac mechanism for patient monitoring in hospitals," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 10, 2018.
- [59] I. Uddin, A. Baig, and A. Ali Minhas, "A controlled environment model for dealing with smart phone addiction," *International Journal of Advanced Computer Science and Applications(IJACSA)*, vol. 9, no. 9, 2018.
- [60] Uddin, "High-level simulation of concurrency operations in microthreaded many-core architectures," *GSTF Journal on Computing (JoC)*, vol. 4, p. 21, 2015.
- [61] Uddin, "One-ipc high-level simulation of microthreaded many-core architectures," *International Journal of High Performance Computing Applications*, vol. 10, 2015.
- [62] Uddin, "Multiple levels of abstractions in the simulation of microthreaded many-core architectures," *Open Journal of Modelling and Simulation*, vol. 3, 2015.
- [63] A. P. Bradley, "The use of the area under the roc curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.