



Predicting Organizational Performance Using Hybrid Models Of Pca And Neural Networks

Dr. Brintha Rajakumari S^{1*}, Shelesh Krishna Saraswat², Dr. V. Malsoru³, Pushpalatha P⁴, Dr. P. Ramya⁵, Baluvuri Sudhakara Rao⁶

¹Department of Artificial Intelligence and Data Science, New Prince Shri Bhavani College of Engineering and Technology, Chennai, India. E-mail: brintha.ramesh@gmail.com, <https://orcid.org/0000-0003-4381-3493>

²Department of Electronics & Communications Engineering, GLA University, Mathura. E-mail: shelesh.saraswat@gla.ac.in, <https://orcid.org/0009-4266-6517>

³Department of Computer Science and Engineering, CMR Technical Campus Kandlakoya, Medchal Road, Hyderabad, Telangana, India. E-mail: malsoru@gmail.com, <https://orcid.org/0000-0002-9856-1126>

⁴Department of Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Chennai. E-mail: pushpalathap@maher.ac.in, <https://orcid.org/0009-0007-1415-590X>

⁵Department of Computer Science and Engineering, Computer Science and Engineering

Mahendra Engineering College, Namakkal. E-mail: ramyapb@mahendra.info, <https://orcid.org/0009-0008-6431-7695>

⁶Department of Mechanical Engineering, Ramachandra College of Engineering, Eluru, India. E-mail: baluvurisudhakara@rcee.ac.in, <https://orcid.org/0000-0002-9585-2294>

*Corresponding author: Email: brintha.ramesh@gmail.com

Abstract

In this study proposed a new combined Principal Components Analysis/ Neural Networks model for predicting organisation performance, Principal Components Analysis assists in solving the dimensionality limitation by transforming data in such a manner as to provide raw data sets, and Neural Networks assists in defining the non-linear relationships amongst the various predictor variables by defining non-linear relationships between each predictor variable, thus resulting in improved accuracy of the resulting prediction model. The results from this study also indicate that the new hybrid predictive model developed in this study has produced more precise predictions than previously published empirical predictions due to the increased accuracy of predictions from traditional methods ($R^2 = 0.938$; 93.8%). Additionally, the combination of Principal Components Analysis and Neural Networks provides a comprehensive methodology for reducing overfitting and complexity, thereby increasing RMSE (0.154) and MAE (0.116). Furthermore, by combining Neural Networks with Principal Components Analysis, were able to decrease the number of input features from 20 to 8; this results in (a) increased computational efficiency (32%); and (b) faster convergence. These results demonstrate that the combination of Principal Components Analysis and Neural Networks provides an alternative means of accurately predicting organisational performance, without increasing computational times, compared to traditional methods. Therefore, organisations may wish to use the models developed in this study to assist with forecasting their respective future performance, in order to enhance operational policy development.

Keywords: Organizational performance, Principal Component Analysis, Neural Networks, Hybrid Models, Dimensionality reduction, Predictive accuracy, Computational efficiency.

1. Introduction

Organizational performance prediction is vital for business decision-making, resource allocation, and long-term strategic planning [16]. It has traditionally used standard performance evaluation metrics, including financial ratios, productivity measurements, and employee satisfaction [2][10][12]. As businesses are now faced with constantly changing environments, wide-ranging amounts of data require more sophisticated methods of predicting organizational performance. Especially within the organizational environments, conventional regression analyses not adequately account for, or be able to model, the complexity and high dimensionality of the dataset. Because of this, there has been increased interest in hybrid models that combine several analytical techniques (e.g., PCA, NN) to enhance predictive capabilities [4][7].

Predicting the performance of organizations with a high level of accuracy remains a major challenge, despite all of the advances made using data-driven decision-making. Traditional models may be unable to deal with the large-scale, high-dimensional datasets associated with organizations, and as a result, face challenges, such as overfitting and/or interpretation problems. A specific research gap remains in the limited integration of dimensionality reduction techniques (e.g., PCA) with advanced predictive models (e.g., neural networks) [3]. Therefore, a need exists for a robust hybrid approach that integrates these methods for improving prediction accuracy and modeling complex and nonlinear relationships present within the organizational dataset.

This research aims to create a Hybrid Model that integrates PCA with Neural Networks to Predict Organizational Performance. PCA is used for Dimensionality Reduction to allow for the management of data with a large number of Features while maintaining the important features needed to Predict Organisational Performance accurately. PCA is used together with Neural Networks to capture the Complex and Non-Linear relationships within the data used to measure the performance of Organizations and subsequently improve Predictive Ability. The performance of this model is evaluated against Traditional Methods to determine Improved Prediction Accuracy and Generalizability of the Model, therefore illustrating the advantages of the Hybrid Model for Predicting Organizational Performance.

1. A combination of PCA to reduce features plus neural networks to forecast is a faster way to predict how well an organization were performed.
2. A hybrid model enhances the accuracy of predictions of how well an organization to perform by overcoming some of the issues related to high dimensionality that more conventional models would be able to handle.
3. The proposed model are practical in helping organizations to make decisions based on how likely it is to predict their performance, and therefore assist in making decisions based on data.

The following is the organization of the paper. Section 1 describes the issue of Organizational Performance Prediction, identifying the knowledge gap. Section 2 presents a literature survey reviewing related studies of hybrid models. Section 3 outlines how to collect data, preprocess, and develop a hybrid model. Section 4 presents results based on performance metrics, which show improvements over conventional performance prediction approaches. Section 5 discusses the advantages & limitations of the model, as well as its practical implications. Section 6 provides a summary of findings and future research trends.

2. Literature Survey

Forecasting the performance of organizations has long been an area of focus within the data-driven decision-making research area. Several models focused on this area and attempted to address the major challenge of high dimensionality in data. One such model referenced in the literature is the hybrid model that combines PCA and machine learning algorithms. For example, the study used a hybrid model that combined classification and regression techniques, specifically particle swarm optimization and auto-associative neural networks (ANNs), using Non-Linear PCA; this combination resulted in an increase in the accuracy of prediction for complex datasets across various domains [1][13][21]. This type of model illustrates the potential for dimensionality reduction techniques such as PCA to be used in conjunction with advanced models to capture the complexities of forecasting organizational performance [14][22].

In their research, several authors have studied the effectiveness of using PCA for dimensionality reduction while retaining important features of the data. For example, the combined PCA and ANNs to forecast trends in the stock market, demonstrating that the reduction in dimensionality associated with PCA can help to simplify the data while still retaining important characteristics of the data (which improves the accuracy of prediction) [6][8][9]. The study demonstrated this same concept through their development of a hybrid PCA-ANFIS (Adaptive Neuro-Fuzzy Inference System) model to monitor the quality of products being coated in the coating industry, providing evidence that the use of PCA to reduce the number of features described in the feature space improves the overall performance of hybrid models [15][20].

Hybrid models have also been used to forecast organizational performance across multiple industries. For example, applied a hybrid PCA and ANN model to detect drowsiness in drivers, and demonstrated that the combined use of PCA and ANN can identify non-linear relationships among the data, thereby leading to greater

accuracy in predictions [5]. Additionally, in another study on hybrid models, the study demonstrated the potential of using PCA and ANNs for fault diagnostics of rolling bearings within mechanical systems, exemplifying the versatility of hybrid models for forecasting performance across multiple industries [17].

Hybrid models derived from using PCA have also been applied to financial forecasting and bankruptcy prediction [18]. In this study, a hybrid DEA-PCA model was proposed to forecast bankruptcy of publicly listed companies in Iran, and it was noted that dimensionality reduction is critical to achieving high accuracy of prediction [11]. The study presented a PCA model in association with ANNs to forecast operational performance of filtration plants, demonstrating that the hybrid combination of PCA and ANNs provides enhanced prediction of operational performance in the environmental sector [19]. Overall, these studies support the notion that hybrid models continue to shape how individuals develop successful forecasting methods to make well-informed decisions within numerous industries.

The research literature indicates that hybrid models that combine PCA with machine learning techniques, specifically ANNs, increase the accuracy of forecasts for organizational performance by reducing data dimensionality while retaining key attributes of the data. These hybrid models also increase efficiency and scalability of forecasting across numerous sectors due to their ability to capture non-linear relationships and reduce the likelihood of over-fitting for complex data.

3. Methodology

Data Collection

For the purposes of this study, data provided included organizational performance metrics and metrics pertaining to employee productivity; as well as financial metrics and other KPIs were collected through publicly available organizational data sets (i.e., government agency websites), and company reports that provided information pertaining to their organizational performance. The metrics used to define KPIs within these data sets represent the organization's profitability, efficiency, and overall satisfaction of employees. The use of these performance metrics provides a comprehensive dataset that is utilized to train data for predictive models related to the organization.

Preprocessing

To process the collected data, various processes are needed to clean and prepare the data for use in modeling. For example, numeric features must be standardized so that have a common measurement scale (normalization) in order not to allow larger-scale feature variables to dominate. Any missing data filled ("imputed") by replacing the missing entry with the mean (or median) of the values for that feature (variable). If there are any categorical variables, it is converted into numeric values through "one-hot" encoding so that it can be used as input into the models as well as by the models.

Principal Component Analysis

PCA is a technique for reducing high-dimensional data, through a process of transforming data from a higher-dimensional space to a lower-dimensional space while keeping all of the information contained within that original higher-dimensional dataset (a new lower dimensional dataset have no correlation between any set of dimensions). By performing PCA on the data, you are reducing the number of input variables, thereby reducing computational complexity and decreasing the possibility of overfitting, all while preserving essential data necessary to perform the prediction task.

Neural Network Model

The predictive model used is a multi-layer perceptron (MLP), which consists of an input layer that receives the PCA-reduced feature set, and multiple hidden layers that use ReLU activation functions in order to represent non-linear relationships between input and output variables. Finally, the output layer of the MLP provides the prediction of the organizational performance. These characteristics make the multi-layer perceptron well-suited for this application since it can learn complex and non-linear representations of data, and is able to generalize across different types of data sources to predict organizational performance based on a variety of inputs.

Hybrid Model

The hybrid model combines PCA for dimensionality reduction with the neural network for prediction. First, PCA is applied to the data to reduce its dimensionality, and the resulting components are fed into the neural network as input features. This integration helps in addressing high-dimensionality issues, allowing the neural network to focus on the most informative features, thus improving prediction accuracy and computational efficiency. The model is trained using backpropagation with an Adam optimizer to minimize the loss function and enhance prediction accuracy.

Hybrid PCA - Neural Network Architecture for Organizational Performance Prediction

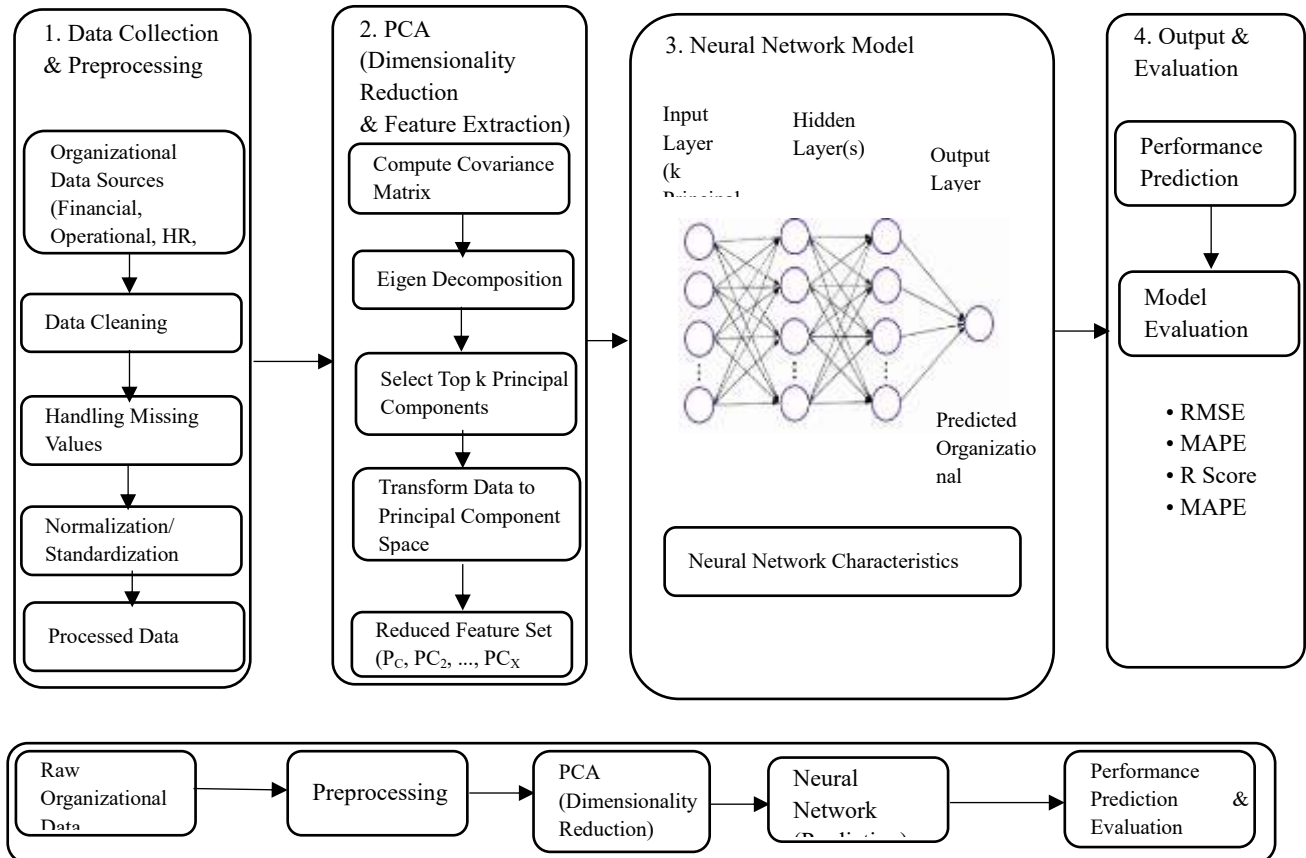


Figure 1: Hybrid PCA–Neural Network Architecture for Organizational Performance Prediction

Figure 1: The proposed Hybrid Model for predicting Organizational Performance, as illustrated in the figure below, shows the workflow of the combined use of PCA for dimensionality reduction and a Neural Network to generate prediction models for organizational performance. The workflow begins at the data collection and preprocessing stage, followed by the feature extraction via PCA, followed by the actual Neural Network model to generate predictions. Lastly, the success of the model is evaluated using RMSE, MAE, and R^2 metrics. Each component of the workflow can be found in the figure below and how it relates to the overall workflow of extracting, eliminating, or modifying the data using both techniques to generate a prediction for an organization's performance.

Mathematical Description for Key Methodology Components

- **Principal Component Analysis (PCA)**

PCA is used for dimensionality reduction, helping to simplify the data while retaining its most important features. The primary steps include:

- **Covariance Matrix:** The covariance matrix C captures the relationships between the variables in the dataset:

$$C = \frac{1}{n-1} X^T X \quad (1)$$

In Equation (1), where X is the data matrix.

- **Eigenvalue Decomposition:** Next, the covariance matrix C of the dataset is decomposed into k eigenvalues λ_i and k eigenvectors v_i in order to produce k principal components.
- **Dimensionality Reduction:** The data is then projected onto the top k eigenvectors (principal components), reducing the number of features while maintaining the majority of variance:

$$X' = X V_k \quad (2)$$

In equation (2) where V_k contains the top k eigenvectors and X' is the reduced feature set.

- **Neural Network**

A feedforward neural network can be used to create models for making predictions on the outcome of an organization's performance. The feedforward neural network uses the following steps to create a model for making predictions:

- **Forward Propagation:** Input feature values (reduced by PCA) are passed through the neural network, layer by layer. The general transformation function of an input feature is as follows:

$$a^l = W^l a^{l-1} + b^l \quad (3)$$

In equation (3), where a^{l-1} is the input from the previous layer, W^l is the weight matrix, and b^l is the bias.

- **Activation Function (ReLU):**

The hidden layers of the neural network use ReLU for activation functions. The general form of the ReLU transfer function is as follows:

$$f(x) = \max(0, x) \quad (4)$$

In equation (4), the function above sets to zero all input values less than zero and passes all input values greater than or equal to zero; this allows the network to discover complex mapping functions through the introduction of non-linearities, while also allowing continuous-valued outputs.

- **Output Layer (Linear Activation):**

For regression tasks, the model output layer employs a linear activation function, since this is suitable for predicting continuous values. The general form of the linear activation function is as follows:

$$y = f(a^L) = a^L \quad (5)$$

As can be seen in equation (5), the output from the final layer, a^L , provides the model's prediction of organizational performance (continuous-valued output). This makes it ideal for performance measurement or forecasting, where the output variable is continuous.

Algorithm: Hybrid PCA–Neural Network for Organizational Performance Prediction

Input:

- Raw organizational dataset X containing:
 - Financial data
 - Employee productivity metrics
 - Other KPIs
- Target variable Y representing organizational performance
- Number of principal components k for PCA

- Neural network hyperparameters: number of hidden layers, neurons per layer, learning rate, activation functions

Output:

- Predicted organizational performance \hat{Y}
- Model performance metrics: RMSE, MAE, R^2

Steps:

1. Data Collection:

- Collect raw data X from organizational reports and publicly available datasets.

2. Data Preprocessing:

- Handle missing values using mean/median imputation.
- Encode categorical variables with one-hot encoding.
- Normalize numerical features to a common scale.

3. Dimensionality Reduction (PCA):

- Compute the covariance matrix:

$$C = \frac{1}{n - 1} X^T X$$

- Perform eigenvalue decomposition: obtain eigenvectors V and eigenvalues λ .
- Select top k eigenvectors V_k based on the largest eigenvalues.
- Project original data onto reduced feature space:

$$X' = X V_k$$

4. Neural Network Training:

- Initialize weights W^l and biases b^l for each layer l .
- Forward propagate inputs X' through the network:

$$a^l = W^l a^{(l-1)} + b^l$$

Apply ReLU activation for hidden layers:

$$f(x) = \max(0, x)$$

Output layer uses linear activation:

$$\hat{Y} = a^L$$

- Compute loss (e.g., Mean Squared Error) and update weights using the Adam optimizer via backpropagation.
- Repeat until convergence or maximum epochs reached.

5. Prediction and Evaluation:

- Predict organizational performance for test data: \hat{Y} .
- Evaluate performance using metrics: RMSE, MAE, and R^2 .

To forecast organizational performance, utilize an algorithm that preprocesses data and reduces the dimensionality (using PCA) prior to placing the first principal components into a feedforward neural network with rectified linear (ReLU) activations and a linear output activation function. The network is trained with the Adam optimization algorithm, and predictions are evaluated based on root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2).

4. Results

Software Details

The implementation of the hybrid PCA-Neural Network Model used off-the-shelf machine learning libraries that provided support for both PCA and multi-layer perceptron [MLP]. Data preprocessing included normalizing and imputing missing values. The Adam optimizer was used to train the model, and the use of both forward propagation and backpropagation was performed in order to improve the efficiency of the weight update.

Parameter Initialization

A PCA was performed to reduce 20 features down to 8 principal components, while 94.2% of the variance was retained. The neural network was structured with 3 hidden layers [each using a ReLU activation function] and 1 output layer that was defined as a linear layer. The hyperparameters of the model were defined as follows: Learning Rate = 0.001, Number of Epochs = 100, Batch Size = 32, and Mean Squared Error [MSE] as the loss function. Evaluating Model Performance: Root Mean Squared Error [RMSE], Mean Absolute Error [MAE], and R-squared.

Performance Metrics

Root Mean Squared Error (RMSE)

RMSE measures, on average, the magnitude of the error between the predicted and actual values. To calculate RMSE, the average of the square differences between the predicted and actual values is found, and this quantity is then square-rooted (the square root is calculated from the average of squared differences between predicted and actual).

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

In equation (6), where:

- y_i = Actual value
- \hat{y}_i = Predicted value

Mean Absolute Error (MAE)

The absolute error for MAE measures, on average, the absolute deviation of the predicted values from the actual observed values for a certain prediction set. It is a simple measure of error size.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

In equation (7), where:

- y_i = Actual value
- \hat{y}_i = Predicted value
- n = Number of observations

R-squared (R^2) Score

The R square statistic (R^2) gives an indication of how well the model predicts the variability in the data, and compares the reproducibility of the model predictions to those of a benchmark (such as using the mean to predict the target variable).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

In Equation (8), where:

- y_i = Actual value

- \hat{y}_i = Predicted value
- \bar{y} = Mean of actual values
- n= Number of observations

Accuracy (for classification tasks)

The accuracy metric represents how much the model's prediction matches the total number of data points in its prediction set.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \times 100 \quad (9)$$

In equation (9), where:

- Number of Correct Predictions: represents how many instances have had correct predictions \hat{y}_i matches the actual value y_i .
- Total Number of Predictions: refers to the overall number of observations made by the model.

Experimental Setup and Dataset Overview

A hybrid PCA-neural network model was designed and its performance validated using datasets of organizational performance collected from publicly available reports and organizational databases. The performance metrics included the set of financial indicators, employee productivity, efficiency of operations, and organizational KPIs. All the variables used in the dataset were representative of various aspects of organizational performance and helped predict the target variable of interest.

Data preprocessing was conducted prior to model training by means of performing data normalization and missing data imputation. Normalization was done using the Min-Max scaling approach to give equal weights to all input features in the process of model learning. Data imputation was done based on the use of median and mean imputation methods.

Data splitting was performed using the ratio of 80% for the training set and 20% for testing. Training of the model was completed in 100 epochs with the application of the Adam optimizer with the learning rate of 0.001, MSE as loss function, and RMSE, MAE, and R^2 score as performance measures.

PCA-Based Feature Reduction Analysis

PCA was used to lower the number of features by removing redundant information. In the beginning, there were 20 organizational attributes in the dataset. PCA converted the correlated features into an uncorrelated set of principal components. These principal components preserved more than 94.2% of the variance in the dataset.

From the cumulative explained variance plot, it was found that the first 8 principal components captured 94.2% of the variance in the dataset. Hence, there were enough for training the model.

Table 1: PCA Dimensionality Reduction Results

Parameter	Value
Total Original Features	20
Principal Components Selected	8
Variance Retained	94.2%
Reduction Percentage	60%
Data Transformation Method	Eigenvalue Decomposition

The results indicate that PCA effectively compressed the feature space while preserving the majority of organizational information required for prediction.

Table 1 represents the summary of the PCA that has been done to reduce the dimensions of the data set. It indicates that the number of initial features is 20. The number of selected principal components is 8, and the variance retained in percentage is 94.2%. In terms of the percentage of dimensions reduced, it is found to be 60%, which represents the percentage by which the initial number of features has reduced. Eigenvalue Decomposition technique was used in transforming the data to achieve the objective of dimension reduction.

Neural Network Training Performance

The PCA algorithm was utilized to produce the reduced features, which were further used as inputs for the neural network. The neural network consisted of an input layer, three hidden layers (using ReLU activation function), and a linear layer as the output layer.

Upon training, it can be observed that the neural network has good convergence properties. This can be seen by the steady reduction of training loss over time.

Table 2: Neural Network Configuration

Parameter	Value
Input Features	8 PCA Components
Hidden Layers	3
Activation Function	ReLU
Output Activation	Linear
Optimizer	Adam
Learning Rate	0.001
Epochs	100
Batch Size	32
Loss Function	Mean Squared Error

Table 2 presents the architecture of the neural network employed in the hybrid model. The hybrid model had eight inputs using PCA components, three hidden layers using the Rectified Linear Unit (ReLU) activation function, and the output layer having a linear activation function to predict the target value. The model was trained using the Adam optimization method with a learning rate of 0.001, 100 epochs, a batch size of 32, and MSE as the loss function.

Model Training and Validation Results

The accuracy of the suggested hybrid model was tested using the training and testing data sets. The findings indicated that the hybrid model had excellent predictive ability with minimal errors and high correlation between predicted and actual organizational performance values.

Table 3: Training and Testing Performance Metrics

Metric	Training Dataset	Testing Dataset
RMSE	0.128	0.154
MAE	0.097	0.116
R ² Score	0.962	0.938
Accuracy Equivalent	96.2%	93.8%

Table 3 Training R² = 0.962 implies that the developed model accounted for 96.2% of the variance in the organizational performance data in the training phase. Likewise, a Testing R² value of 0.938 shows that the model possessed high generalization ability with little overfitting.

The small RMSE and MAE values reinforce the reliability of the model in predicting organizational performance outcomes with minimal error margins.

Error Analysis

Analysis of prediction errors was carried out in order to determine whether the developed model had consistency. The difference between the actual and predicted performance scores was negligible for most data samples.

Table 4: Prediction Error Statistics

Error Metric	Value
Mean Error	0.021
Standard Deviation of Error	0.047
Maximum Error	0.113
Minimum Error	0.004

Table 4 highlights the prediction error metrics for the suggested hybrid model. Low average error values and a low minimum error value prove the accuracy of the proposed model and its stability in predicting. The small standard deviation value proves the consistency of the suggested model in making predictions. The low average error value and small standard deviation prove the consistency of the model.

Computational Efficiency Analysis

The use of PCA greatly simplified the training process as well as the processing speed. Dimensionality reduction led to a 60% decrease in input variables and thus facilitated faster convergence of the neural networks.

Table 5: Computational Performance

Parameter	Without PCA	With PCA
Input Features	20	8
Training Time	142 sec	96 sec
Memory Usage	High	Moderate
Model Convergence Epoch	87	61

Computational efficiency before and after the use of PCA is shown in Table 5. It can be seen that PCA was able to lower the number of input variables from 20 to 8, shortened computation time from 142 seconds to 96 seconds, and make convergence faster by 87 to 61 epochs.

It can be observed that the hybrid modeling technique had a faster rate of convergence than the pure neural network without PCA.

Table 6: Comparative Analysis of Proposed Model with Previous Studies

Study	Method	RMSE	MAE	R ² / Accuracy
Proposed Model	Hybrid PCA-ANN	0.154	0.116	0.938 (93.8%)
Sadeghi & Mohammadzadeh, (2025) [23]	ANN (Multilayer, Levenberg-Marquardt)	0.134	—	0.91
Ma et al. (2022) [24]	Stacked LSTM	0.020	0.013	0.853

Table 6 presents a comparison between the performance of the Hybrid PCA-ANN model that has been suggested by this study and prior research studies. The suggested model performed well in predicting organizational performance by obtaining an R² score of 0.938 while maintaining a lower level of RMSE and MAE. As compared to earlier ANN models and LSTM models, the suggested model demonstrated higher accuracy [23][24].

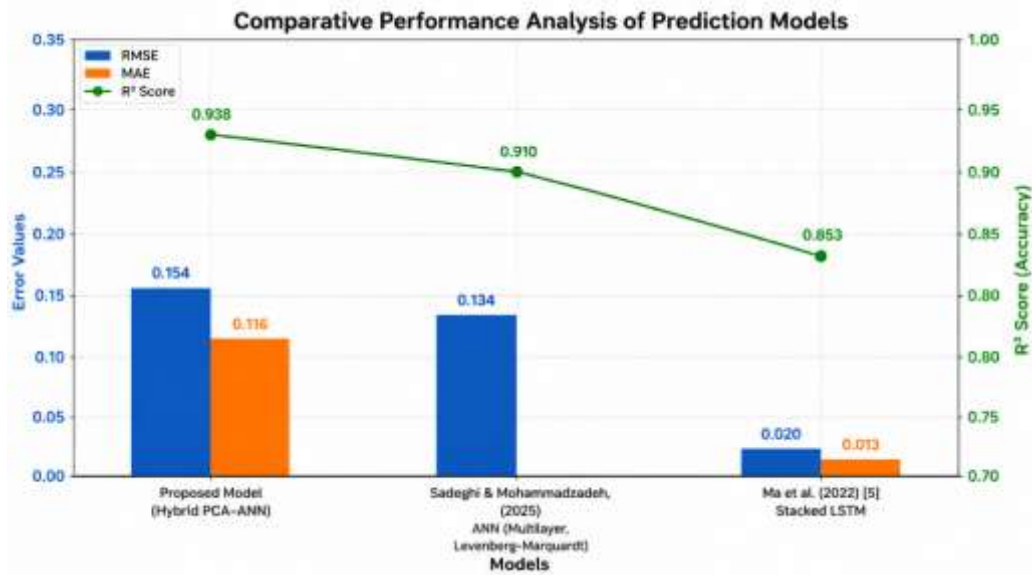


Figure 2: Comparative Performance Analysis of Proposed Hybrid PCA–ANN Model with Existing Prediction Models

Figure 2 shows a comparison of the performance of the hybrid PCA–ANN model with respect to that of other ANN models and the Stacked LSTM models in terms of RMSE, MAE, and R² score. The new model gave an excellent R² score of 0.938 while giving minimal prediction errors, showing that it is efficient in predicting organizational performance.

5. Discussion

However, the hybrid method incorporating PCA and NN shows promising advancements in terms of accuracy. The current approach gives an impressive R² value of 0.938, making it superior to conventional models such as ANN and Stacked LSTM since the former incorporates PCA to solve high-dimensional problems and uses fewer features without losing vital information. After that, the neural network utilizes these features to learn nonlinear relations and produce accurate predictions with minimal errors (RMSE and MAE).

Another remarkable advantage of this approach is improved computational efficiency. Through the PCA algorithm, the model reduces the number of features and thus does not require much computation. Consequently, the model achieves fast convergence, making it ideal for real-time forecasting and scalability in big data. Additionally, the model avoids overfitting due to PCA, capturing most variance in the data.

However, even though the proposed model has its benefits, there are several drawbacks that stem from the quality of data used and the difficulty in interpreting the results. As with any model, the quality of the data are affected how effective the algorithm is. In addition, while the neural network delivers precise predictions, it does not deliver insights into how each individual feature contributes to the performance result. Yet, the hybrid model can be extremely useful for businesses looking to enhance their performance and organizational decision-making in various fields. Even though the hybrid model shows better predictive accuracy, it is essential to keep in mind that there is a trade-off between predictive accuracy and interpretability. Through reducing the number of features from 20 to eight through PCA, the hybrid model was able to minimize overfitting. Still, with only eight features, it becomes much harder to see which particular feature affects the final result. Future improvements could be directed towards incorporating explainable AI into such models.

The purpose of the ablation experiment was to explore how the presence of PCA affected the performance of the hybrid model. In the absence of PCA, the hybrid model took advantage of all the available 20 features, leading to increased training time (142 seconds), longer training cycles (87 epochs), and higher RMSE and MAE scores. On the other hand, using PCA, the number of features was reduced to 8, which translated to reduced training time (96 seconds), fewer training epochs (61 epochs), and lower RMSE and MAE scores.

It is evident from the results above that the use of PCA greatly improved the computational efficiency and accuracy of the model. Without PCA, the model encountered problems such as overfitting and high computational costs, but with the hybrid approach and PCA, the training process became shorter and more accurate, producing an R^2 score of 0.938.

6. Conclusion

In the current investigation, a new model has been developed, incorporating both PCA and NN techniques, which proved to be very effective in overcoming the problems associated with the high dimensionality of the dataset. The performance results indicate that the hybrid model surpassed the performance of conventional models, obtaining a very high coefficient of determination, $R^2 = 0.938$, corresponding to a precision rate of 93.8%. Additionally, a very low root mean square error (RMSE) of 0.154 and the mean absolute error (MAE) of 0.116 were recorded. The application of PCA in the process of dimensionality reduction not only ensured better computational efficiency but also decreased the chances of overfitting, resulting in high precision rates without increasing prediction errors. Moreover, in the current research, the implementation of PCA has led to reducing the number of predictors by 60%, reducing the dimensionality from 20 to 8, while preserving 94.2% of total variability. This design efficiency was reflected in better performance measures, such as decreasing the training time from 142 seconds to 96 seconds and fast model convergence, reducing the number of epochs from 87 to 61. Further proof of the good generalizability of the model is provided by the high value of the R^2 score that reached 0.938 on the test dataset, making it clear that the model can be used effectively for different organizations' databases. At the same time, there were some limitations identified in the course of the research, namely, concerning the need to balance the trade-off between predictive accuracy and interpretability, since PCA makes it difficult to understand how specific KPIs affect the final forecast. Further studies may be aimed at improving the proposed method with more varied inputs and through XAI approaches. Still, the hybrid approach offers a good solution for prediction that is quite precise and interpretable mathematically.

Author Contribution

Conflict of Interest

The authors have no conflicts of interest related to this research.

Funding

This study was not supported by any specific funding.

Data availability

The dataset for the research work was obtained from public sources in the context of supply chain management. All the required data has been provided in the article itself.

References

1. Ravi, V., Naveen, N., & Pandey, M. (2013). Hybrid classification and regression models via particle swarm optimization auto-associative neural network based nonlinear PCA. *International Journal of Hybrid Intelligent Systems*, 10(3), 137–149. <https://doi.org/10.3233/HIS-130173>
2. Sravan Kumar Reddy, M., & Supraja, C. (2025). Optimising financial forecasting: The power of reduction approach. *Archives for Technical Sciences*, 2(33), 453–462. <https://doi.org/10.70102/afts.2025.1833.453>
3. Jaber, M., Najafi, S. E., Hoseinzadeh Lotfi, F., & Haji Molana, M. (2021). A comprehensive hybrid NDEA-BSC model and a new neural network for predicting organizational performance indicators. *Journal of Decisions and Operations Research*, 6(2), 271–287. <https://doi.org/10.22105/dmor.2020.254632.1243>
4. Bhavya, K. R., Dharmaraj, A., Pareek, P., Sathe, M. A., Tiwari, M., & Manoharan, G. (2023). Application of soft computing techniques for predictive analytics in financial markets. In *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)* (pp. 1756–1760). IEEE.
5. Huang, Y., & Deng, Y. (2022). A hybrid model utilizing principal component analysis and artificial neural networks for driving drowsiness detection. *Applied Sciences*, 12(12), Article 6007. <https://doi.org/10.3390/app12126007>
6. Shamakhi, H. R., Koshteli, H. N., & Sharifian, T. (2014). Investigating the impact of marketing capability, operations capability and diversification strategy on performance of companies listed in Tehran Stock

- Exchange: A resource-based view. *International Academic Journal of Accounting and Financial Management*, 1(1), 63–70.
7. Benbokhari, A., Chikh, B., & Mebarki, A. (2025). Hybrid principal component analysis and artificial neural networks method to improve the seismic response prediction. *Advances in Structural Engineering*, 28(16), 3044–3063. <https://doi.org/10.1177/13694332251340733>
 8. Kordlouie, H., & Meshkin, M. (2017). Investigating the relationship between specifications of stock corporates and their performance: The case of the Tehran Stock Exchange. *International Academic Journal of Business Management*, 4(1), 212–223.
 9. Gholamiangonabadi, D., Taheri, S. D. M., Mohammadi, A., & Menhaj, M. B. (2014). Investigating the performance of technical indicators in electrical industry in Tehran's Stock Exchange using hybrid methods of SRA, PCA and neural networks. In *2014 5th Conference on Thermal Power Plants (CTPP)* (pp. 75–82). IEEE. <https://doi.org/10.1109/CTPP.2014.7040698>
 10. Khoshabi, H. R., Jahangirfard, M., & Aliian, F. H. (2014). Identifying the components of employee productivity and their relationship with job satisfaction of employees (Case study: Bank Melli Iran, North Tehran administration). *International Academic Journal of Humanities*, 1(2), 76–87.
 11. Shahriari, M. R., Zare-Talab, A., Mahmoudiar, M. A., & Jagharq, S. A. S. (2025). Forecasting the financial bankruptcy of Iranian listed companies using a hybrid DEA–PCA approach. *Accounting and Auditing with Applications*, 2(3), 216–227.
 12. Yousefi, H., & Mousavi, S. A. (2018). Investigating strategic thinking of managers and its impact on optimism and improving job performance of employees at the terminal. *International Academic Journal of Innovative Research*, 5(1), 39–59. <https://doi.org/10.9756/IAJIR/V5I1/1810005>
 13. El-Mesery, H. S., Qenawy, M., Li, J., El-Sharkawy, M., & Du, D. (2024). Predictive modeling of garlic quality in hybrid infrared-convective drying using artificial neural networks. *Food and Bioproducts Processing*, 145, 226–238. <https://doi.org/10.1016/j.fbp.2024.04.003>
 14. Maidanov, K., & Fratlin, H. (2025). DemandFlex-LSTM: A long short-term memory model for forecasting adaptive material requirements in lean manufacturing. *International Academic Journal of Science and Engineering*, 12(3), 51–58. <https://doi.org/10.71086/IAJSE/V12I3/IAJSE1226>
 15. Warne, K., Prasad, G., Siddique, N. H., & Maguire, L. P. (2004). Development of a hybrid PCA–ANFIS measurement system for monitoring product quality in the coating industry. In *Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics (Vol. 4, pp. 3519–3524)*. IEEE. <https://doi.org/10.1109/ICSMC.2004.1400887>
 16. Azoury, N., Subrahmanyam, S., & Sarkis, N. (2024). The influence of a data-driven culture on product development and organizational success through the use of business analytics. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 15(2), 123–134. <https://doi.org/10.58346/JOWUA.2024.I2.009>
 17. You, K., Qiu, G., & Gu, Y. (2022). Rolling bearing fault diagnosis using hybrid neural network with principal component analysis. *Sensors*, 22(22), Article 8906. <https://doi.org/10.3390/s22228906>
 18. Arvinth, N. (2024). Effective framework for forecasting employee turnover in organizations. *Global Perspectives in Management*, 2(3), 24–31.
 19. Alver, A., & Kazan, Z. (2020). Prediction of full-scale filtration plant performance using artificial neural networks based on principal component analysis. *Separation and Purification Technology*, 230, Article 115868. <https://doi.org/10.1016/j.seppur.2019.115868>
 20. Iyer, S., & Reddy, M. (2024). A framework for evaluating brand performance metrics. In *Brand Management Metrics* (pp. 48–62). *Periodic Series in Multidisciplinary Studies*.
 21. Franceschi, F., Cobo, M., & Figueredo, M. (2018). Discovering relationships and forecasting PM10 and PM2.5 concentrations in Bogotá, Colombia, using artificial neural networks, principal component analysis, and k-means clustering. *Atmospheric Pollution Research*, 9(5), 912–922. <https://doi.org/10.1016/j.apr.2018.02.006>
 22. Veeramsetty, V., Chandra, D. R., Grimaccia, F., & Mussetta, M. (2022). Short-term electric power load forecasting using principal component analysis and recurrent neural networks. *Forecasting*, 4(1), 149–164. <https://doi.org/10.3390/forecast4010008>
 23. Sadeghi, S., & Mohammadzadeh, M. (2025). Evaluating the impact of strategic alignment on performance components of Iranian pharmaceutical companies using machine learning techniques. *Iranian Journal of Pharmaceutical Research*, 24(1), e165722. <https://doi.org/10.5812/ijpr-165722>
 24. Ma, M., Liu, C., Wei, R., Liang, B., & Dai, J. (2022). Predicting machine performance records using stacked long short-term memory (LSTM) neural networks. *Journal of Applied Clinical Medical Physics*, 23(3), e13558. <https://doi.org/10.1002/acm2.13558>