KeAi CHINESE ROOTS GLOBAL IMPACT
International Journal of Innovation Studies

IOT AWARE AUTOENCODER DEEP CAPSULE NETWORK FOR EFFICIENT PREDICTIVE ANALYTICS IN CLOUD

Mrs. M. Nandhiya[1], Dr. C.R Durga devi[2]
[1]PhD Research Scholar, Department of Computer Science, NGM College, Pollachi, Tamil Nadu, India. Email: nandhuya25@gmail.com,
nandhiya.m@ngmc.org.
[2]Assistant Professor, Department of Information Technology, NGM College, Pollachi, Tamil Nadu, India. Email: deviswe@gmail.com, durgadevi@ngmc.org.

Abstract

Cloud computing is a term used to portray the delivery of on-demand computing resources such as servers, storage, databases, software, and analytics through the internet. Cloud computing facilitates organizations to access and store information without handling their own physical devices. In cloud computing model, predictive analytics is the process of forecasting the future outcomes or trends using historical data and machine learning algorithms. Many existing models have achieved lesser accuracy in predicative analytics model and increased their complexity due to variations in key features across different regions. A novel **P**olynomial **R**egressive **R**obust **A**utoencoded **D**eep **C**apsule Network (PRRA-CapsNet) model is developed for efficient predictive analytics in cloud computing. The main aim of PRRA-CapsNet model is to improve the accuracy of prediction with minimum time. The proposed **d**eep **c**apsule network includes the five different layers, namely one input layer, one output layer and three hidden layers namely convolutional layer, primary capsule network and class capsule network. Initially, IoT sensors are employed to gather the data points from different location. After that, the collected data points are transmitted to the input layer. The input layer transmitted the data points to convolutional layer. In the convolutional layer, PRRA-CapsNet model performs data preprocessing which includes missing data handling and outlier removal. Then, the pre-processed data points are transmitted to the primary capsule network. In that layer, Robust Autoencoder Feature Selection is employed in PRRA-CapsNet model for addressing the dimensionality reduction issues to identify the most important features in a dataset. The relevant features are transmitted to the class capsule network. In that layer, the Kulczynski similarity function is used in PRRA-CapsNet model for performing the data classification in cloud environment. This in turn, efficient data analytics is carried out in cloud environment. Experimental analysis is carried out with the performance metrics like classification accuracy, classification time, precision, recall, f-measure and false positive rate with respect to number of data points.

Keywords: IoT, cloud computing, predictive analytics, **D**eep **C**apsule Network, Robust Autoencoder Feature Selection, Kulczynski similarity function

1. Introduction

A cloud computing setting is a digitally controlled system that delivers the various computing services such as data storage, processing capabilities, networking, and software applications through the internet. This computing model offers advantages such as on-demand scalability, operational flexibility, compact infrastructure costs, and efficient system management. An IoT-enabled cloud computing environment expands this conception by integrating Internet of Things (IoT) technologies to gather continuous real-time data from their surroundings and broadcast it to cloud servers. The cloud server then collect, stores, processes, and analyzes this large volume of data by means of advanced computing technology. By combining IoT with cloud computing supports a wide range of real time applications, including smart farming, digital healthcare systems, environmental monitoring, smart city development and so on.

A novel hybrid learning model called Transformer_Gated Recurrent Unit (TRANS_GRU) was developed in [1] with the aim of improving the accuracy of prediction and decreasing the error rate. However, the performance of the TRANS_GRU hybrid model did not show improvement, mainly when it was applied to a larger dataset. A Multiscale Spatial-Temporal-Variable Feature Fusion Network (MSTVFFN) was introduced in [2] aimed for predicting multiple air pollutants. However, the model's sensitivity performance did not improve, since it did not undergo hyperparameter tuning.

To enhance the long-term prediction accuracy, advanced data analysis and deep learning methods were developed in [3]. However, it failed to capture more complex spatiotemporal dependencies, thereby reducing its predictive accuracy. A hybrid forecasting framework was developed in [4] for air pollutant concentrations prediction by using Bidirectional Long Short-Term Memory (BiLSTM) networks. However, the generalizability and scalability of the framework was not improved. Physics-Based Machine Learning (PBML) model was introduced in [5] for offering more accurate and interpretable air pollution predictions. However, the model did not provide sufficiently accurate predictions to efficiently direct air quality management strategies.

An integration of Kalman filtering technique and seasonal gated recurrent units (SGRU) were developed in [6] for predicting air pollution levels. However, achieving high detection accuracy and convenient data acquisition were major issues. A deep learning-based Long Short-Term Memory (LSTM) model was introduced in [7] for significantly increases the air pollution prediction. However, the computational complexity of the model was not reduced. Gaussian Process Regression (GPR) machine learning model was developed in [8] for more efficient air pollution control and environmental sustainability. However, the model did not applying formal statistical tests to evaluate the significance of performance across models. A Conditional Wasserstein Generative Adversarial Network was developed in [9] for accurate prediction based on spatiotemporal features. However, it failed to provide an innovative technology for prediction based on spatiotemporal multi-scale feature extraction. An artificial neural network approach was introduced in [10] for air pollution prediction. Though the model captured more complex relationships and long-term dependencies, performance was limited by the absence of hyperparameter settings.

Bidirectional Temporal Convolutional Network and Enhanced Informer model was introduced in [11] for accurate air pollutant concentration prediction. However, long-term dependence features, and spatial features was not extracted to increase the model's prediction accuracy. An integration of Convolutional Neural Network-Bidirectional Long Short-Term

Memory (CNN-BiLSTM) architecture was developed in [12] for efficiently forecasting atmospheric pollutants. However, enhancing the model's generalization capability and prediction accuracy was major issues. Convolutional Neural Network (CNN) and a Quantum Long Short-Term Memory (QLSTM) network model were developed in [13] for air pollutants prediction with high robustness and efficiency. However, hybrid learning approaches failed to further reduces the time complexity of prediction. Deep Sparse Transformer Networks was developed in [14] for predicting pollutant time series based on long-term dependencies. However, the efficiency of the network was not improved. Data-driven approach based on Deep Learning (DL) and Long Short-Term Memory (LSTM) algorithm were developed in [15] for forecasting the air pollutants. However, the proposed approach exhibited high prediction errors.

1.1 Proposal key contribution

The limitations of existing methods are overcome by introduction of a novel PRRA-CapsNet model. The key contributions of the PRRA-CapsNet model are summarized as follows:

- ❖ To propose a novel PRRA-CapsNet model for efficient predictive analytics in cloud by implementing deep capsule network which integrates different processing steps namely, data pre-processing, feature selection, classification and fine tuning.
- ❖ To minimize the prediction time, PRRA-CapsNet model performs two major processes namely data preprocessing and feature selection. Polynomial predictive mean imputation method is employed for handling the missing data within the dataset. The Modified Z-Score method is employed to detect and remove the outlier data. Moreover, the robust autoencoder is employed for relevant feature selection. These processes of PRRA-CapsNet model receive minimal time for prediction.
- ❖ To enhance the accuracy, Kulczynski similarity function is employed for analyzing the training and testing data samples and provides the accurate prediction results. Besides, horse herd optimization algorithm is designed to fine tune the hyperparameter to enhance the accuracy by reducing the error.
- ❖ In conclusion, complete experimentation is carried out to analyze the performance of our PRRA-CapsNet model and other deep learning framework.

1.2 paper Organization

The rest of this manuscript is organized into different sections as follows. Section 2 provides a widespread analysis of previous studies and existing research. Section 3 outlines the proposed PRRA-CapsNet model, detailing its overall framework and architectural design. Section 4 describes the experimental arrangement and offers a comprehensive analysis. Section 5 presents the experimental result and delivers an in-depth relative assessment of the proposed method against existing deep learning techniques using numerous performance metrics. Finally, Section 6 summarizes the major conclusions and highlights the key contributions of the paper.

2. Related works

Deep aggregation seq2seq network was introduced in [16] for air pollutant concentration prediction. However, the approach was high computational complexity. A new Spatio-temporal inverted Transformer (ST-iTransformer) model was developed in [17] to improve air pollution forecast accuracy. However, the model's flexibility was not improved in air pollution forecasting. A novel spatiotemporal graph attention network (STGATN) was introduced in [18] with an encoder-decoder architecture for predicting air pollutant concentrations. However, it failed to enhance the robustness and generalizability of the model. Relief- Based Feature Selection and ANN model were developed in [19] for air pollution prediction. However, it did not focus on incorporating the fine tuning process to enhance the predictive ability of the model. Deep Neural Networks (DNNs) and Convolutional Neural Networks (CNNs) were developed in [20] for predicting air pollution concentrations based on multivariate time series data. However, the model failed to improve its adaptability and general applicability.

Ensemble machine learning model was introduced in [21] for reliable air pollution prediction. However, accurate and timely air quality forecasting remained major issues. Shallow and deep sparse autoencoder artificial neural network was developed in [22] to enhance the air pollution index prediction. However, the model required longer training time. Echo State Network (ESN) deep learning model was introduced in [23] to improve the accuracy of air pollution prediction. However it failed to explore new effective metaheuristic algorithms for hyperparameter optimization. Residual-enhanced hybrid forecasting framework was developed in [24] for Air pollution forecasting. However, it failed to perform the large-scale environmental monitoring applications. Machine and deep learning methods were introduced in [25] for air pollution prediction using meteorological variables. However, efficient hybrid approaches were not implemented to further improve the prediction accuracy.

Multilayer Perceptron (MLP) network was developed in [26] for increasing the accuracy of urban air pollution modeling. However, significant reduction in air pollution time was major issues. An integrated artificial intelligence (AI) framework was developed in [27] to predict, forecast and analyze major air pollutants. However, multi-pollutant AQI metrics was not accurately predicted. A Nonlinear Autoregressive network with Exogenous inputs (NARX) model was introduced in [28] to predict concentrations of key pollutants. However, it did not analyze the temporal and spatial features to improve predictive accuracy. LSTM neural networks were developed in [29] for accurate air pollution forecasting. However, the model did not efficient for handling large scale dataset. Hybrid Clustering Large Applications (CLARA) with the Fuzzy Time Series Markov Chain (FTSMC) approach were developed in [30] for reliable air pollution forecasting. However, achieving high accuracy and robustness of the predictions were major issues.

3. Proposal methodology

Predictive analytics is the utilization of data, statistical algorithms, and machine learning techniques to classify the probability of future outcomes depends on historical events. In this paper, a novel deep learning model called PRRA-CapsNet model has been developed for accurate predictive analytics with minimal time consumption by implementing the cloud based infrastructure. A deep capsule network is an advanced type of neural network to overcome some key issues of traditional neural networks.

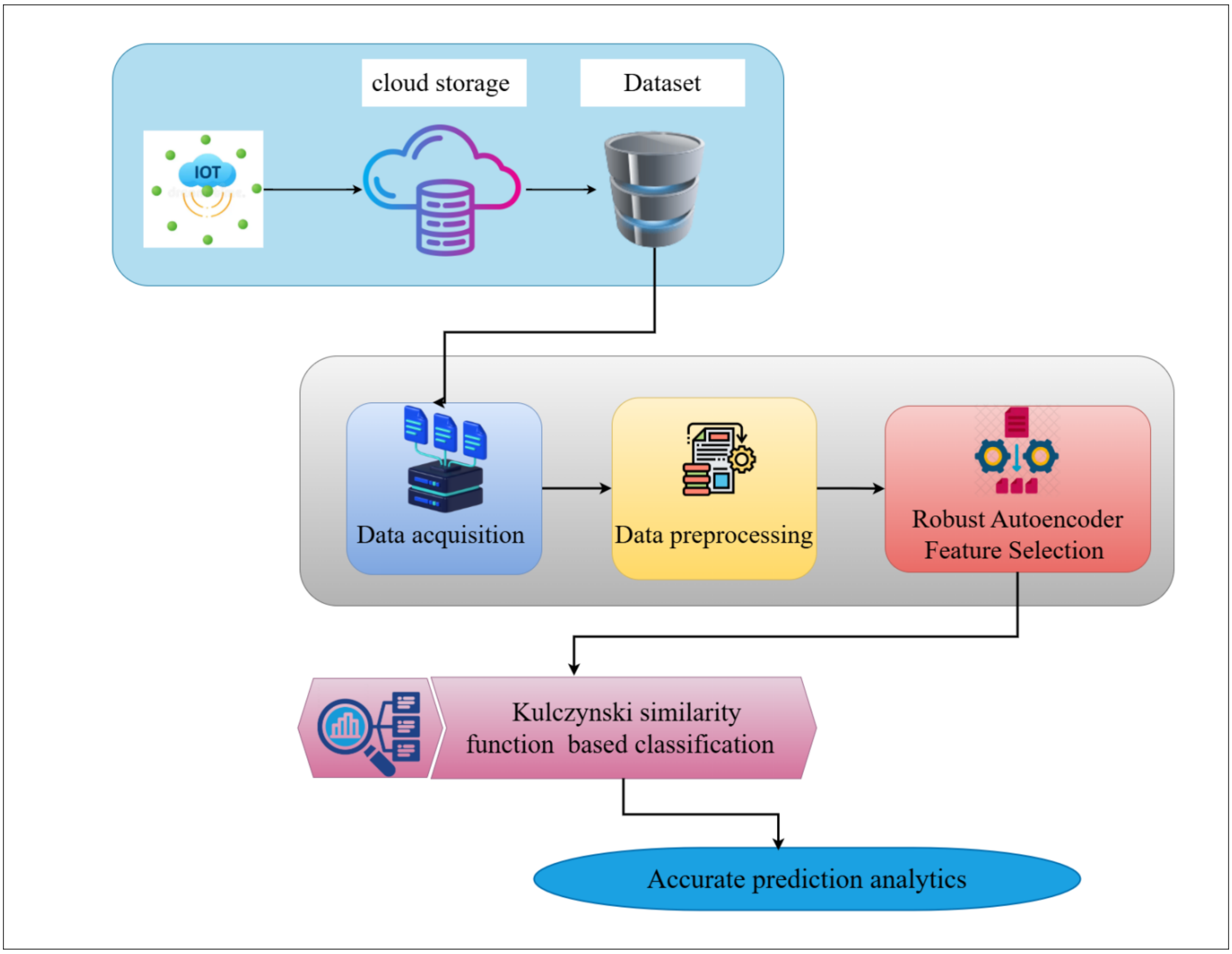

Figure 1. Architecture of the proposed PRRA-CapsNet model

Figure 1 illustrates the comprehensive architecture of the proposed PRRA-CapsNet model, developed to deliver accurate predictions within a cloud-based environment. The proposed model is organized into numerous sequential stages, namely data acquisition, preprocessing, feature selection, and classification. To begin with, a substantial volume of data is gathered for reliable predictive modeling. This dataset provides the major input to the system. The overall architecture analytically integrates preprocessing, feature selection, and classification into an integrated network. In the preprocessing stage, raw data undergo transformation to convert it into a structured format for accurate predicative analysis. Following preprocessing stage, relevant feature selection is performed using an auto encoder based approach. This method identifies the most influential attributes while removing irrelevant or redundant features, reducing computational complexity and improving model accuracy. By retaining only the most informative features, the proposed model ensures improved predictive performance. To end with, the refined feature set is transferred to the classification module, where the model executes the predictive analytics task. This process considerably enhances the prediction accuracy and system performance. The subsequent sections provide a comprehensive explanation of each stage within the proposed architecture.

3.1 Data acquisition

Data acquisition is the foundational processing stage of the proposed PRRA-CapsNet model and plays an essential role in ensuring consistent predictive outcomes in cloud based environments. In this framework, IoT-enabled sensors and intelligent monitoring systems are developed. These devices communicate by means of wireless networking technologies, enabling smooth and continuous transmission of collected data to a centralized cloud platform. The cloud infrastructure stores, and manages the incoming data, making it actively available for advanced processing and analytical tasks. In addition to sensor data, the PRRA-CapsNet model also incorporates information obtained from publicly accessible datasets. By integrating multiple data streams, the PRRA-CapsNet enhances the model's ability to produce accurate, meaningful, and reliable predictions.

3.2 Deep capsule network based predictive analytics in cloud computing

A deep capsule network is a type of neural network that performs accurate predictive analytics. The main advantage of deep capsule network is to achieve the good accuracy with minimal training time. The architecture of CapsNet is partitioned into three main parts. These parts are the input layer, hidden layer, and output layer. The capsule networks utilize the groups of neurons called capsules that produce the output vectors.

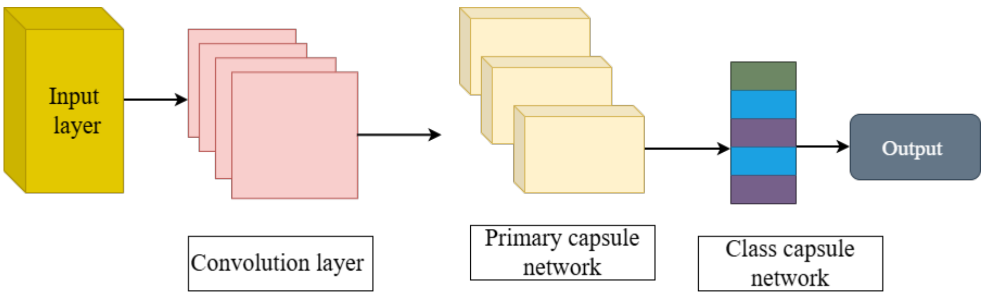

Figure 2 Schematic structure of deep CapsNet

Figure 2 given above depicts the schematic structure of a deep CapsNet network classifier, which includes three different types of layers such as an input layer, one or more hidden (middle) layers, and a one output layer. The input and output layers are typically single layers, while the hidden layers consist of three sublayers, such as convolutional layers, primary capsule layer, and class capsule layer, which integrated together to analyze the input data samples. Each layer contains numerous small processing units known as artificial neurons or called capsules that produce the output vectors. These capsules receive input, perform certain computations tasks, and transfer the processed output to the next layer. Finally, the output layer produces and final prediction results.

Let us consider the input dataset 'DS' which includes a sample of data samples 'SR' as well as attributes or features $\{A_1, A_2, \dots, A_m\}$ are arranged in the form of matrix. Therefore, the dataset samples and their related attributes are organized into an input matrix as formulated below.

$$X = \begin{bmatrix} A_1 & A_2 & \dots & A_{\mathrm{m}} \\ SR_{11} & SR_{12} & \dots & SR_{1n} \\ SR_{21} & SR_{22} & \dots & SR_{2n} \\ \vdots & \vdots & \dots & \vdots \\ SR_{m1} & SR_{m2} & \dots & SR_{mn} \end{bmatrix} \quad (1)$$

Where, X represents an input matrix where each column indicates a number of features $A = \{A_1, A_2, \ldots, A_m\}$, each row represents a number of data samples 'SR' = $\{SR_1, SR_2, \ldots, SR_n\}$' respectively

These input matrixes are given to the input layer. In that input layer, each neuron received the input and it transferred into the convolution layer. The neurons in the layer compute the weighted sum of the input and the bias.

$$Z = \sum(\mathrm{X} * W_{ih}) + B \quad (1)$$

Where, Z denotes an activity of the neuron in convolutional layer, W_{ih} represents a weights between input and hidden layer, input matrix 'X' , 'B' indicates a bias that stored the value is '1'.

- **Convolutional layer**

In the hidden layer 1 (Convolutional layer), data preprocessing is performed to prepares raw data for efficient analysis and accurate prediction. Preprocessing step is essential to enhance efficiency, improve result accuracy, and reduce the time. The preprocessing method includes two different processes namely missing data handling and outlier removal. A missing value is a value that designates that no data value is stored for the particular feature in the current observation. Missing values in a dataset is identified and resolved by employing the polynomial predictive mean imputation method. In this imputation method, first polynomial regression is employed to find the missing value with the available data points. Therefore, the regression process is expressed as follows,

$$SR_{Miss} = \varphi_0 + \varphi_1 SR_1 + \varphi_2 {SR_2}^2 + \varphi_3 {SR_3}^2 \ldots \varphi_n {SR_n}^n (2)$$

Where, SR_{Miss} indicates a missing data value, $\varphi_0, \varphi_1, \varphi_2 \ldots \varphi_n$ indicates a polynomial coefficients, which are determined based on available data points, $SR_1, SR_2, \ldots, SR_n$ indicates known data samples in the dataset. The estimated missing data values 'SR_{Miss} ' are imputed into the null entries, efficiently handling missing data. Followed by, the mean is computed with the estimated missing values.

$$\mu = \frac{1}{n} \sum_{i=1}^{n} SR_i \quad (3)$$

Where, μ denotes a mean value, n indicates a number of data samples 'RS'. After that, the distance between its predicted value and each observed mean is computed as follows,

$$D = \left|\mu_{observed} - \mu_{predicted}\right| (4)$$

Where, D denotes a distance, $\mu_{predicted}$ denotes a predicted mean values using (4), $\mu_{observed}$ denotes a observed mean values. Followed by, minimal distance is determined for finding the most similar to the missing case in terms of the model's predicted outcome.

$$w = \arg\min D \quad (5)$$

Where, w denotes an outcomes of imputation, $\arg\min D$ denotes an argument of minimal distance. The predicted cases with minimal distance are the most similar to the missing case in terms of the model's predicted outcome.

Followed by, the outlier's are determined data points that differ considerably from the majority of other data samples in a dataset. The model utilizes Modified Z-Score method for detecting the outlier's data from the dataset.

To begin with, the input data samples for specific features are organizing in an ascending order. After that, the median value 'SR_{Med}' is computed in the ascending order. Subsequently, the statistical test is expressed as follows,

$$MZS = \vartheta \left(\frac{|SR_i - SR_{Med}|}{dev_{MA}} \right) \quad (6)$$

Where, MZS indicates a Modified Z-Score outlier test analysis, SR_i represents a data in the particular cell, SR_{Med} represents a median value of particular samples, dev_{MA} denotes a mean absolute deviation, $|SR_i - SR_{Med}|$ represents a absolute deviation between the data and their median value, ϑ denotes a scaling factor and the value is 0.6745 to match standard deviation units. The MZS test output ranges from 0 to 1.

$$Q = \begin{cases} MZS > Th, & Outlier \\ MZS \leq Th, & not\ outlier \end{cases} \quad (7)$$

If the estimated MZS value is greater than the threshold 'Th', the particular data sample is considered an outlier. Otherwise, the estimated MZS value is lesser than the threshold 'Th', the data is considered an outlier. In this process, the entire outlier data points within the dataset are identified. Once detected, these outliers are eliminated from the dataset to ensure that subsequent predictions are accurate and reliable. The preprocessed results are transferred to the primary capsule network for selecting the significant features from the dataset.

- **Primary capsule network layer**

After the data preprocessing, the primary capsule network is a key model in capsule networks, which are a higher type of neural network designed to recognize spatial hierarchies and relationships in data. In this network layer, spatial relationships between the features are determined. Feature selection is the process of choosing significant features for reliable, accurate prediction. The proposed model utilizes the robust autoencoder is a type of deep neural network that comprises of two functions such as encoder and decoder. The encoder function is used to transform the input raw data into a lower-dimensional representation, whereas the decoder function restructures the original input dataset from this encoded representation.

A Robust Autoencoder (RAE) provides numerous advantages over traditional autoencoders. RAEs are designed to deliver more reliable results when the data is imperfect and automatically identify irrelevant features, enhancing overall data quality. This process produces a refined version of the original dataset, making it more suitable for analysis.

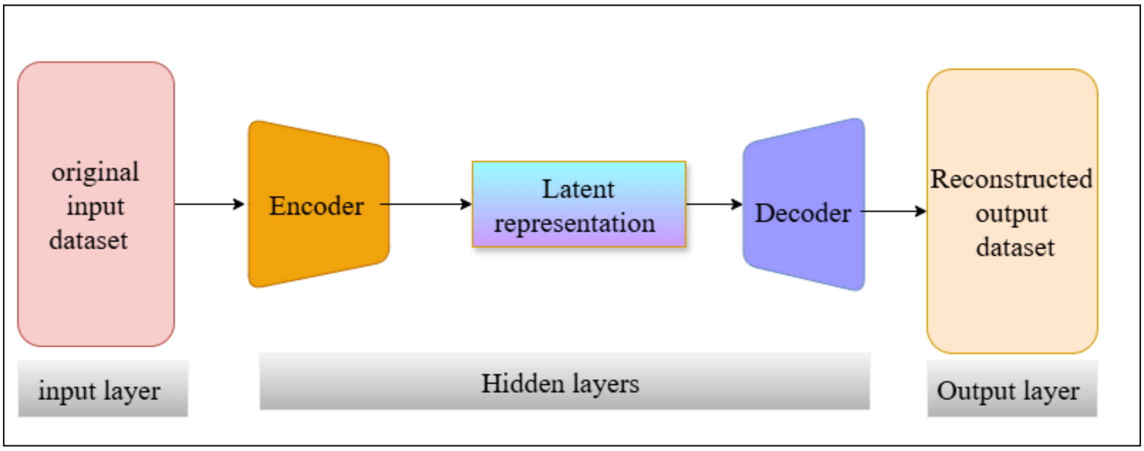

Figure 3 network structure of Robust Autoencoder

Figure 3 demonstrates the schematic structure of a Robust Autoencoder which comprises of two major functions such as encoder and decoder. The input preprocessed dataset is given to the input layer of autoencoder. Then the input preprocessed dataset is then transferred into the hidden layers which consist of two layers namely encoder and decoder. The encoder and decoder layers are used for processing the given input data samples.

In the encoder layer, this model utilizes Azadkia–Chatterjee's dependence coefficient to discover and maintain significant features while discarding the unrelated ones.

Let us consider the features $\{A_1, A_2, \ldots, A_m\}$ in high-dimensional space gathered from the dataset.

$$ACD = 1 - \frac{\sum_{j=1}^{m}\sum_{k=1}^{c}|A_j - T_k|}{ST} \quad (8)$$

Where, ACDdenotes a Azadkia–Chatterjee's dependence coefficients, T_k indicates a 'k' target output, ST indicates wald statistic test statistics,

$$ST = \frac{(A_j - T_k)^2}{Var(A_j, T_k)} \quad (9)$$

Where $Var(A_j, T_k)$ indicates variance between the features and target samples. The correlation ' ACD'returns the output ranges from '0 to '1'.

$$ACD = \begin{cases} 1, & Relevant\ features \\ 0, & Irrelevant features \end{cases} \quad (10)$$

If the outcome of 'ACD' is more close to 1, specifies relevant feature, whereas 0 indicates an unrelated features.

The latent space depiction of an autoencoder is a reduced, lower-dimensional representation of the input by the encoder part of the network. It captures the most significant patterns in the data while removing irrelevant data.

$$LR = \vartheta\left(w_E A_j + b_E\right) \quad (11)$$

Where, LR indicates a latent representation or encoded output of the input dataset, A denotes an input feature vector to the encoder, w_E indicates a weight matrix of the encoder layer. It transforms the input feature vector into the latent space, b_E indicates a bias vector of the encoder layer, ϑ denotes an activation function to introduce nonlinearity, permitting the network to find out irrelevant patterns. Therefore, the decoder maps the latent vector 'LR' back to the original input space.

$$G = \vartheta(w_D LR + b_D) \quad (12)$$

Where, G indicates a decoded output, LR indicates a latent space representation, w_D indicates a weight matrix of the decoder layer, b_D indicates a bias vector of the decoder layer, b_D indicates an activation function. The primary objective of a robust autoencoder is to minimize the reconstruction loss within the compressed, low-dimensional representation of the data. This objective is formulated mathematically as follows,

$$arg\,min\left|A_j - G\right| + \lambda|K| \quad (13)$$

Where, A_j indicates an input data matrix to a standard deep autoencoder, G indicates a reconstructed output of the autoencoder, λ indicates a regularization parameter that manages the sparsity level, S indicates a sparse matrix to separate irrelevant features, $arg\,min$ indicates a argument of minimal function. Based on the analysis, relevant features are preserved for further processing, while irrelevant ones are removed to improve the accuracy. The selected features are then passed to the third hidden layer i.e. class capsule layer.

- **Class capsule layer**

After selecting the more significant features, the next process focuses on accurately perform data analytics in cloud computing. Classification is performed in class capsule layer by analyzing the relationship between the training and testing data samples using Kulczynski similarity coefficient. Let us consider the training data samples with selected features RS_r. In order to perform the data analytics, Kulczynski similarity function is employed for analyzing the testing and training data samples.

For each input samples, the relationship score between the testing and training data samples is measured by applying a Kulczynski similarity function to measure the relationship between the pixels.

$$KS = \frac{[SR_r \cap SR_t]}{[SR_r \cup SR_t]+[SR_r \triangle SR_t]} \quad (14)$$

Where, KS indicates a Kulczynski similarity function offers the relationship between the training samples SR_r and the testing data samples 'SR_t', $SR_r \cap SR_t$ denotes a mutual dependence between the samples, $SR_r \triangle SR_t$ represents a symmetric divergence between the samples, $SR_r \cup SR_t$ indicates the mutual independence between data samples. The similarity function provides the output values from 0 to 1 $[0 \leq KS \leq 1]$.

This similarity-analysis enhances the robustness and accuracy of the predictive analytics.

For each prediction outcome, the error is calculated depends on squared variation between the actual and prediction results.

$$PR = [Y_{Act} - Y_{pred}]^2 \quad (15)$$

Where, PR denotes an error, Y_{Act} represents actual results, Y_{pred} represents predicted results. In order to minimize the classification error of capsule network, the fine tuning process is employed by employing Nadam method

$$W_{new} = W_t - \eta F_t \quad (16)$$

$$F_t = \gamma F_{t-1} + (1-\gamma)\left(\frac{\partial PR}{\partial W_t}\right) \quad (17)$$

Where, W_{new} symbolizes an updated new weights between the layers, W_t represents a previous weight, η represents a learning rate, '$\frac{\partial PR}{\partial W_t}$' represents a partial derivative of the prediction error 'PR' regarding current weight 'W_t', γ indicates default constant value 0.9. Through the updating process, various weight values are observed. From the updating weight value, optimal one is chosen by implementing the wild horse herd optimizer.

From the updating process, numerous weights are observed based on the computed gradients. Therefore, an optimal weight is determined by applying wild horse herd optimizer toward the accuracy enhancement by minimizing the prediction error (i.e. false positive or false negative). Wild horse herd optimizer is a metaheuristic optimization algorithm to mimic the social dynamics and movement patterns monitored in their herds efficiently.

By employing this optimization algorithm, the horses are related to the weights. To begin with, the population of the horses (i.e. number of weights 'W_b') are initialized within search space.

$$W_b = W_1, W_2, W_3 \ldots . W_h \quad (18)$$

After the initialization, the fitness is computed depends on prediction error.

$$fit(W_b) = \arg\min PR \quad (19)$$

Where $fit(W_b)$designates a fitness of each weight value, $arg\ min$ designates an argument of minimal function, PRindicates an error rate in prediction. After evaluating the fitness values, the best-performing horse, representing the optimal weight set, is identified from the population. This horse is assigned as the leader of the herd. Once the leader is chosen, the subsequent subsections explain various behavioral approaches of the horses, including exploration and exploitation methods used during the optimization process.

- **Exploration phase**

During the exploration phase of the optimization process, the horses (search agents) move randomly and diversely to recognize competent new solutions. This stochastic exploration helps prevent early convergence and facilitates the algorithm to thoroughly examine the search space.

$$P_{t+1} = P_t + r_1 * 0.5|P_{best} - P_t| \quad (20)$$

Where,P_{t+1} symbolizes an updated position of the horse, P_t indicates a current position of the horse, r_1 denotes a random number ranged from 0 to 1, $0.5|P_{best} - P_t|$ symbolizes a divergence function between the current position of horse 'P_t' and best position of the horse 'P_{best}'.

- **Exploitation phase**

During the exploitation phase, each horse in the herd updates its position by moving toward the best-known solution. In this manner, random movements are reduced to permit a more focused and accurate local search around the optimal region.

$$P_{t+1} = P_{best} + r_2 * 0.5|P_{best} - P_t| \quad (21)$$

Where, P_{t+1} designates updated position, P_{best} symbolizes current best location of the horse, r_2 indicates a random number (0,1), $0.5|P_{best} - P_t|$ represents a divergence, the current position 'P_t and 'P_{best}' indicates a best position of horse. After updating the positions of the horses during either the exploration or exploitation phase, the algorithm reevaluates their fitness values to assess the quality of the new solutions

$$H = \begin{cases} fit(P_{t+1}) > fit\ (P_{best}) \ ; & P_{t+1}\ is\ optimal \\ Otherwise \ ; & P_{best}\ is\ optimal \end{cases} \quad (22)$$

Where, H indicates a fitness verification results, if the recently updated position 'P_{t+1}' has a better fitness value than the best 'P_{best}', it returned newly updated position P_{t+1} as an optimal. Otherwise the best 'P_{best}' is chosen as optimal. The process continues iteratively until the predetermined maximum number of iterations is reached. Finally, the optimal position of the horse represents the best set of weights is identified for minimizing the error in prediction. Finally, the multi class prediction output is observed at output layer of capsule network architecture with softmax activation function.

$$Y = f_{softmax}(C(t)) \quad (23)$$

Where, Y' represents a final multi class prediction output, '$C(t)$' represents the Class capsule layer output, ' $f_{softmax}$' symbolizes the softmax activation function to provides the multi class prediction results at the output layer. The algorithmic process of **p**olynomial **r**egressive **r**obust **a**utoencoded **d**eep **c**apsule network model is given below.

//Algorithm 1: Polynomial Regressive Robust Autoencoded Deep Capsule Network
Input: preprocessed dataset 'DS', selected features $A_1, A_2, \dots, A_b$, data samples $SR' = \{SR_1, SR_2, \dots, SR_n\}$ **Output:** Improve the prediction accuracy
Begin **// Data acquisition** **Step 1: Collect n**umber of data samples data samples $SR' = \{SR_1, SR_2, \dots, SR_n\}$ **a**nd features $A_1, A_2, \dots, A_b$ from the dataset **// Data processing** **Step 2: Input features** $A_1, A_2, \dots, A_m$ and data samples $SR_1, SR_2, \dots, SR_n$ given to input layer **Step 3:** Transform input to Convolutional layer **Step 4: For each** data samples SR_i **do** **Step 5**: Compute the neuron activation probability using (1) **Step 6: End For** **// Data Pre-processing ------ Convolutional layer** **Step 7: For each** data samples SR_i **do** **Step 8:** Perform missing data imputation using (2) (3) (4) (5) **Step 9: End for** **Step 10: For each data** samples SR_i **do** **Step 11: Compute** Modified Z-Score outlier test 'MZS using (6) **Step 12: If** $(MZS > Th)$ **then** **Step 13:** Detect outlier data samples **Step 14: else** **Step 15:** Detect normal data samples **Step 16: End if** **Step 17: End for** **// feature selection ------ Primary capsule network layer** **Step 18: For each preprocessed** data samples **do** **Step 19**: Perform feature selection using robust autoencoder **Step 20: for each features---- Encoder** **Step 21: Compute the** Azadkia–Chatterjee's dependence coefficient using (8) **Step 22:** $\boldsymbol{if}$ $(ACD = 1)$ **then** **Step 23:** Features said to be relevant **Step 24: elseif** $(ACD = 0)$ **then** **Step 25:** Features said to be irrelevant **Step 26**: **End if** **Step 27: End for** **Step 28: For each encoded output** **Step 29:** Obtain reconstruction output at decoder using (12) **Step 30: End for** **Step 31:** Obtain the reconstructed output at output layer using (13) **// classification------ class capsule layer**

Step 32: **For each selected** feature with training data samples
Step 33: **For each testing** data samples
Step 34: Measure the Kulczynski similarity function ' KS'using (14)
Step 35: $\boldsymbol{if}$ $(KS = 1)$ **then**
Step 36: Data samples classified into particular class
Step 37: **End if**
Step 38: **End for**
Step 39: **End for**
// Fine tuning
Step 40: **For each classification result**
Step 41: Compute the error 'PRusing (15)
Step 42: Update the weight using (16) (17)
Step 43: **Initialize the population of** weights using (18)
Step 44: **For each** W_b
Step 45: Compute the fitness using (19)
Step 46: Identify current best weight with better fitness
Step 47: **While** $(\boldsymbol{t} < t_{max})$ **do**
Step 48: **for each** weight in population **do**
Step 49: **if exploration phase then**
Step 50: Update the positions ' P_{t+1}' according to equation (20)
Step 51: **Else if exploitation phase then**
Step 52: Update the positions P_{t+1}' according to equation (21)
Step 53: **End if**
Step 54: End **for**
Step 55: **Verify fitness of updated** positions using (22)
Step 56: **if** $(fit(P_{t+1}) > fit\ (P_{best})\ )$ **then**
Step 57: P_{t+1} is said to be best optimal solution
Step 58: **else**
Step 59: P_{best} considered as optimal solution
Step 60: **End if**
Step 61: Increment **t= t+1**
Step 62: **Go to step 47**
Step 63: **End while**
Step 64: **Return** optimal weight
Step 65: Obtain final classification results at output layer using (23)**------(output layer)**
End

Algorithm 1 given above illustrates a various process involved for predictive analytics in cloud computing environment. First, the numbers of data samples are gathered from the dataset. Then the received input data samples are given to the input layer of proposed capsule network architecture. After that, it transferred from input to convolutional layer where the neuron activation probability in computed based on weights and biases. Then the data samples undergo preprocessing for solving the missing data and outlier removal. In the next primary capsule network layer, significant features are chosen by employing the robust autoencoder model from the dataset to decrease the dimensionality of the dataset with the help of Azadkia–Chatterjee's dependence coefficient. Finally, classification is carried out in class capsule layer by means of the selected relevant features using Kulczynski similarity function, which calculates the similarity between the training and testing samples. After the classification, error rate is minimized by introducing the fine tuning process using the wild horse herd optimization algorithm. To begin with, a population of horses (i.e. weights) is initialized. Followed by, the fitness is computed depends on classification error rate. With the estimated fitness, the positions of each horse get updated iteratively until reaches maximum iterations. As a final point, the accurate prediction results are attained at the output layer by applying the Softmax function, which further enhances the accuracy with minimal error.

4. Experimental Setup

In this sub-section, experimental assessment of PRRA-CapsNet model and existing TRANS_GRU [1], MSTVFFN [2], are implemented in python high level programming language using Global Air Pollution Dataset taken from https://www.kaggle.com/datasets/hasibalmuzdadid/global-air-pollution-dataset . The primary objective of this dataset is to assess air pollution levels using Air Quality Index (AQI) values derived from numerous pollutants across various cities worldwide. Air pollution is defined as the pollution of indoor or outdoor air by chemical, physical, or biological agents that modify its natural characteristics. Typical sources of air pollutants include domestic combustion systems, vehicular emissions, industrial activities, and wildfires. Key pollutants that create serious risks to public health include particulate matter (PM), carbon monoxide (CO), ozone (O_3), nitrogen dioxide (NO_2), and sulfur dioxide (SO_2). Exposure to both indoor and outdoor air pollution is associated with respiratory diseases and other health complications, making it a leading contributor to morbidity and premature mortality globally. The dataset employed in this study contains 23,463 instances with 12 attributes. Table 1 provides a detailed description of the features used for air pollution prediction.

Table 1 attribute information

S. No	Features	Description
1	Country	Name of the country
2	City	Name of the city
3	AQI Value	Overall AQI value of the city
4	AQI Category	Overall AQI category of the city
5	CO AQI Value	AQI value of Carbon Monoxide
6	CO AQI Category	QI category of Carbon Monoxide
7	Ozone AQI Value	AQI value of Ozone
8	Ozone AQI Category	AQI category of Ozone of the city
9	NO2 AQI Value	AQI value of Nitrogen Dioxide
10	NO2 AQI Category	AQI category of Nitrogen Dioxide
11	PM2.5 AQI Value	QI value of Particulate Matter with a diameter of 2.5 micrometers or less of the city
12	PM2.5 AQI Category	QI category of Particulate Matter with a diameter of 2.5 micrometers or less of the city

4.1 Implementation results

The proposed PRRA-CapsNet model is extensively analyzed to determine its performance in predictive analytics applications. The evaluation process consists of numerous stages, including data acquisition, data preprocessing, feature selection, and classification. For this analysis, the Global Air Pollution Dataset is taken from the Kaggle repository. In the initial phase, a comprehensive set of data samples is gathered from the dataset, as illustrated in Figure 4.

Global Air Pollution Data Acquisition

```
ï»¿Country,City,AQI Value,AQI Category,CO AQI Value,CO AQI Category,Ozone AQI Value,Ozone AQI Category,NO2 AQI Value,NO2 AQI Category,PM
Russian Federation,Praskoveya,51,Moderate,1,Good,36,Good,0,Good,51,Moderate
Brazil,Presidente Dutra,41,Good,1,Good,5,Good,1,Good,41,Good
Italy,Priolo Gargallo,66,Moderate,1,Good,39,Good,2,Good,66,Moderate
Poland,Przasnysz,34,Good,1,Good,34,Good,0,Good,20,Good
France,Punaauia,22,Good,0,Good,22,Good,0,Good,6,Good
United States of America,Punta Gorda,54,Moderate,1,Good,14,Good,11,Good,54,Moderate
Germany,Puttlingen,62,Moderate,1,Good,35,Good,3,Good,62,Moderate
Belgium,Puurs,64,Moderate,1,Good,29,Good,7,Good,64,Moderate
Russian Federation,Pyatigorsk,54,Moderate,1,Good,41,Good,1,Good,54,Moderate
Egypt,Qalyub,142,Unhealthy for Sensitive Groups,3,Good,89,Moderate,9,Good,142,Unhealthy for Sensitive Groups
China,Qinzhou,68,Moderate,2,Good,68,Moderate,1,Good,58,Moderate
Netherlands,Raalte,41,Good,1,Good,24,Good,6,Good,41,Good
India,Radaur,158,Unhealthy,3,Good,139,Unhealthy for Sensitive Groups,1,Good,158,Unhealthy
Pakistan,Radhan,158,Unhealthy,1,Good,50,Good,1,Good,158,Unhealthy
Republic of North Macedonia,Radovis,83,Moderate,1,Good,46,Good,0,Good,83,Moderate
France,Raismes,59,Moderate,1,Good,30,Good,4,Good,59,Moderate
India,Rajgir,154,Unhealthy,3,Good,100,Unhealthy for Sensitive Groups,2,Good,154,Unhealthy
Italy,Ramacca,55,Moderate,1,Good,47,Good,0,Good,55,Moderate
United States of America,Phoenix,72,Moderate,1,Good,4,Good,23,Good,72,Moderate
```

Figure 4 data acquisition

After completing the data acquisition phase, preprocessing procedures were performed to enhance data quality. These procedures involved treating missing data and identifying outlier values. Once the missing data were addressed, the dataset size measured approximately 1,593KB. Following the removal of outliers, the dataset was further refined, resulting in a reduced size of 1,341KB. The outcomes of these preprocessing steps are presented in Figure 5.

Data Preprocessing

```
Country,City,AQI Value,AQI Category,CO AQI Value,CO AQI Category,Ozone AQI Value,Ozone AQI Category,NO2 AQI Value,NO2 AQI Category,PM2.5 AQI V
Russian Federation,Praskoveya,51,Moderate,1,Good,36,Good,0,Good,51,Moderate
Brazil,Presidente Dutra,41,Good,1,Good,5,Good,1,Good,41,Good
Italy,Priolo Gargallo,66,Moderate,1,Good,39,Good,2,Good,66,Moderate
Poland,Przasnysz,34,Good,1,Good,34,Good,0,Good,20,Good
France,Punaauia,22,Good,0,Good,22,Good,0,Good,6,Good
Germany,Puttlingen,62,Moderate,1,Good,35,Good,3,Good,62,Moderate
Belgium,Puurs,64,Moderate,1,Good,29,Good,7,Good,64,Moderate
Russian Federation,Pyatigorsk,54,Moderate,1,Good,41,Good,1,Good,54,Moderate
Egypt,Qalyub,142,Unhealthy for Sensitive Groups,3,Good,89,Moderate,9,Good,142,Unhealthy for Sensitive Groups
China,Qinzhou,68,Moderate,2,Good,68,Moderate,1,Good,58,Moderate
Netherlands,Raalte,41,Good,1,Good,24,Good,6,Good,41,Good
Pakistan,Radhan,158,Unhealthy,1,Good,50,Good,1,Good,158,Unhealthy
Republic of North Macedonia,Radovis,83,Moderate,1,Good,46,Good,0,Good,83,Moderate
France,Raismes,59,Moderate,1,Good,30,Good,4,Good,59,Moderate
India,Rajgir,154,Unhealthy,3,Good,100,Unhealthy for Sensitive Groups,2,Good,154,Unhealthy
Italy,Ramacca,55,Moderate,1,Good,47,Good,0,Good,55,Moderate
India,Phulabani,161,Unhealthy,2,Good,71,Moderate,0,Good,161,Unhealthy
Poland,Piaseczno,28,Good,1,Good,28,Good,2,Good,28,Good
India,Pimpri,118,Unhealthy for Sensitive Groups,2,Good,30,Good,2,Good,118,Unhealthy for Sensitive Groups
```

Figure 5 after data preprocessing

Following the preprocessing stage, a feature selection process is performed to reduce the dimensionality of the dataset. The model employs the robust autoencoder to evaluate relationships among variables and identify the most informative attributes. By analyzing the computed similarity scores, less significant features are eliminated while the most relevant ones are retained. Based on these score values, the model selects seven essential features that contribute significantly to improving the accuracy of prediction.

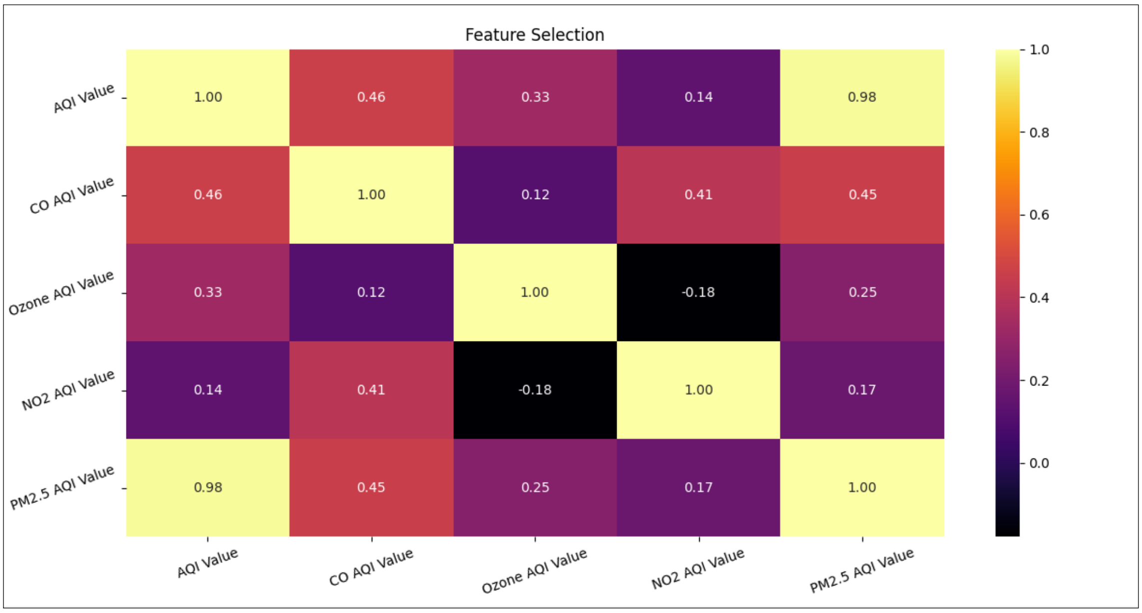

Figure 6 heatmap for correlation between features

Figure 7 selected features

Figure 7 illustrates the feature selection outcomes which identify seven relevant attributes to perform air pollution prediction.

In the final phase, the model carries out air pollution prediction by applying Kulczynski similarity function implemented into the deep capsule network. Using the seven selected features, the PRRA-CapsNet model is accurately predicting the air pollution based on air quality index is good, moderate and unhealthy, very unhealthy and hazardous. The categories Unhealthy, Very Unhealthy, and Hazardous indicate that the air pollution levels are high and potentially harmful to health.

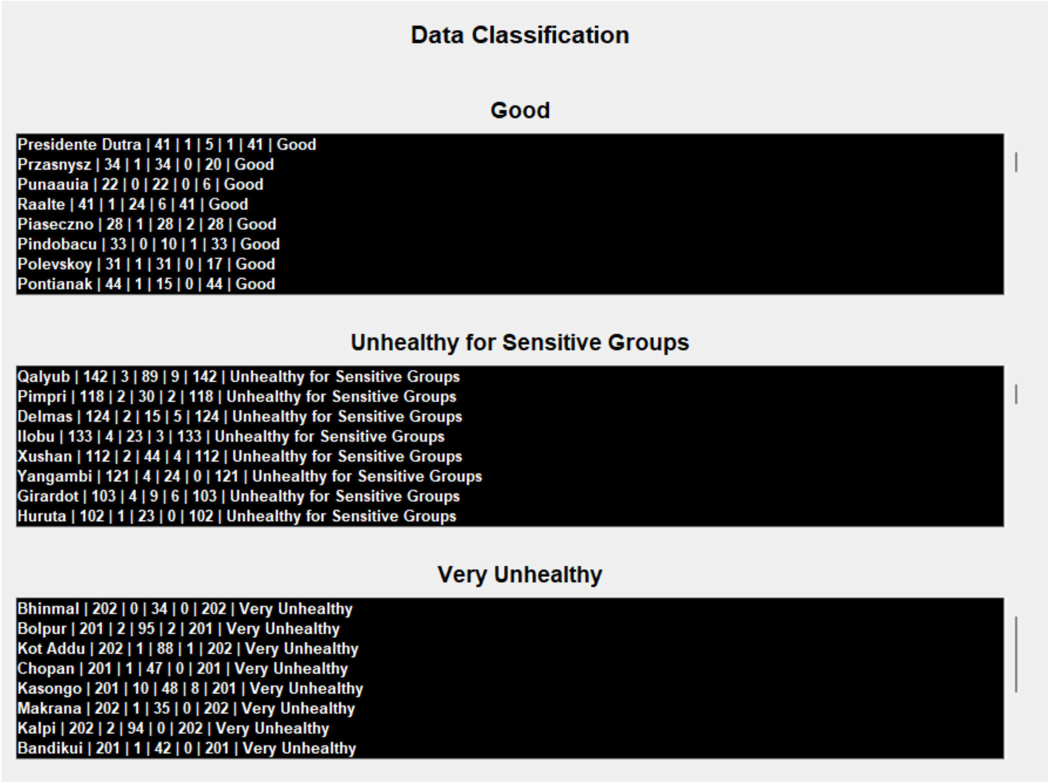

Figure 8 Prediction outcomes

5. Performance comparison analysis

In this **sub**section, performance analysis of the proposed PRRA-CapsNet model existing TRANS_GRU [1], MSTVFFN [2], are analyzed using various metrics, such as prediction accuracy, precision, recall, F1 score, root mean square error (RMSE), Mean absolute error and prediction time.

5.1 Performance comparison of prediction accuracy

The accuracy is defined as the ratio of correctly predicted data samples to the total number of data samples taken from the dataset. The prediction accuracy is mathematically calculated as follows.

$$PA = \frac{TP+TN}{TP+TN+FP+FN} * 100 \quad (24)$$

Where, PA represeents prediction accuracy**,** true positive (TP), true negative (TP)**,** the false positive (FP), false negative 'FN'**.** It is measured in terms of percentage (%).

Table 2 Prediction accuracy

Number of data samples	Prediction accuracy (%)		
	Proposed PRRA-CapsNet	Existing TRANS_GRU [1]	Existing MSTVFFN [2]
2000	96.75	92.5	93.5
4000	96.37	92.65	93.42
6000	96.55	92.15	93.36
8000	96.25	92.35	94.52
10000	96.44	92.45	94.22

12000	96.28	92.35	93.63
14000	96.17	92.42	94.21
16000	96.85	92.35	94.22
18000	96.42	92.4	94.3
20000	96.33	92.47	93.63

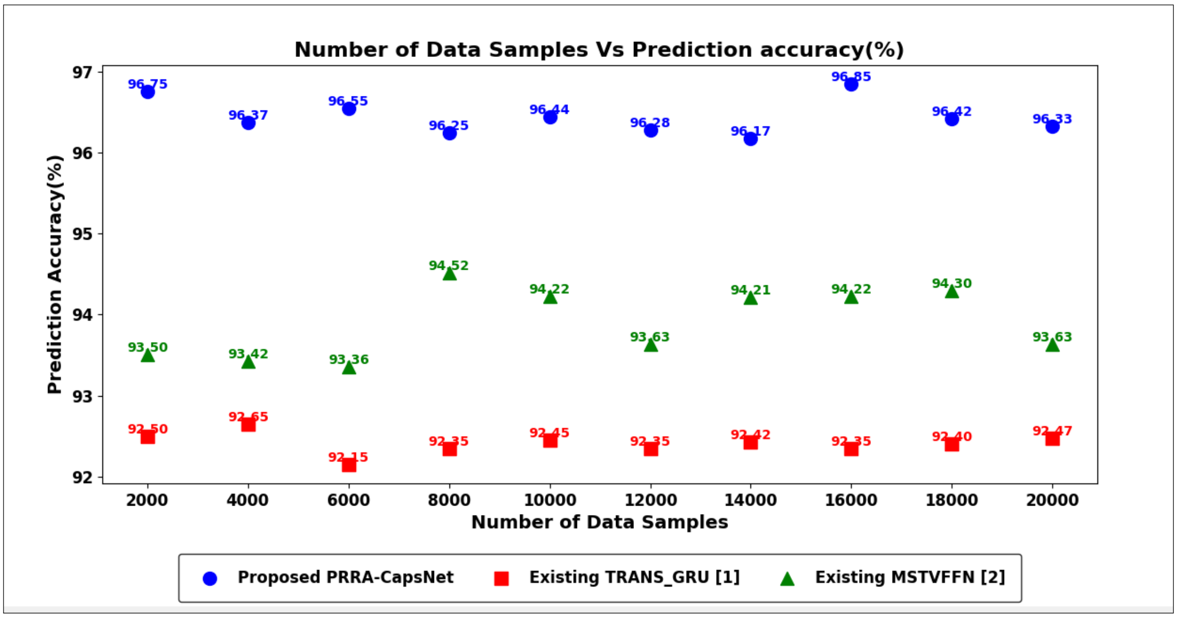

Figure 9 performance analysis of prediction accuracy

Figure 9 shows the prediction accuracy against number of data samples varying from 2000 to 20000, as collected from the dataset. As shown in figure 9, the horizontal axis symbolizes the number of sample data, whereas the vertical axis reveals the equivalent prediction accuracy. The graph demonstrates that the proposed PRRA-CapsNet model outperforms the existing approaches TRANS_GRU [1], MSTVFFN [2]. Let us consider the input data samples of 2000 in the first iteration. By applying the PRRA-CapsNet model, enhanced prediction accuracy was found to be 96.75%. In comparison, the accuracy of the existing methods [1] and [2] were found to be 92.5% and 93.5%, respectively. For each method, ten various results were observed and compared. The average of these ten assessment results demonstrate that the prediction accuracy of the PRRA-CapsNet model improved considerably by 4% and 3% when compared to existing methods [1] and [2], respectively. This superior performance is attained owing to the application of a Deep Capsule Network with the help of Kulczynski similarity function for analyzing training and testing data samples within the class capsule layers. By analyzing these samples, it efficiently improved accuracy in prediction accuracy. Moreover, the fine-tuning process of the proposed Deep Capsule Network utilizes wild horse herd optimization algorithm to reduce classification error by enhancing the true positive and true negative rates, while minimizing the false positives and false negatives in the more accurate pollution prediction.

5.2 Performance comparison of precision

It is an accurate metric used to calculate the accuracy of a model's positive predictions. It is defined as the ratio of true positives (TP) to the total number of predicted positive samples, which consists of both true positives (TP) and false positives (FP).

$$Precision = \frac{TP}{TP+FP} \quad (25)$$

Where, true positive rate (TP), false positive rate (FP).

Table 3 precision

Number of data samples	precision		
	Proposed PRRA-CapsNet	Existing TRANS_GRU [1]	Existing MSTVFFN [2]
2000	0.976	0.951	0.957
4000	0.975	0.95	0.956
6000	0.974	0.948	0.958
8000	0.973	0.946	0.957
10000	0.972	0.945	0.956
12000	0.977	0.946	0.958
14000	0.976	0.944	0.953
16000	0.975	0.946	0.954
18000	0.973	0.945	0.955
20000	0.972	0.946	0.953

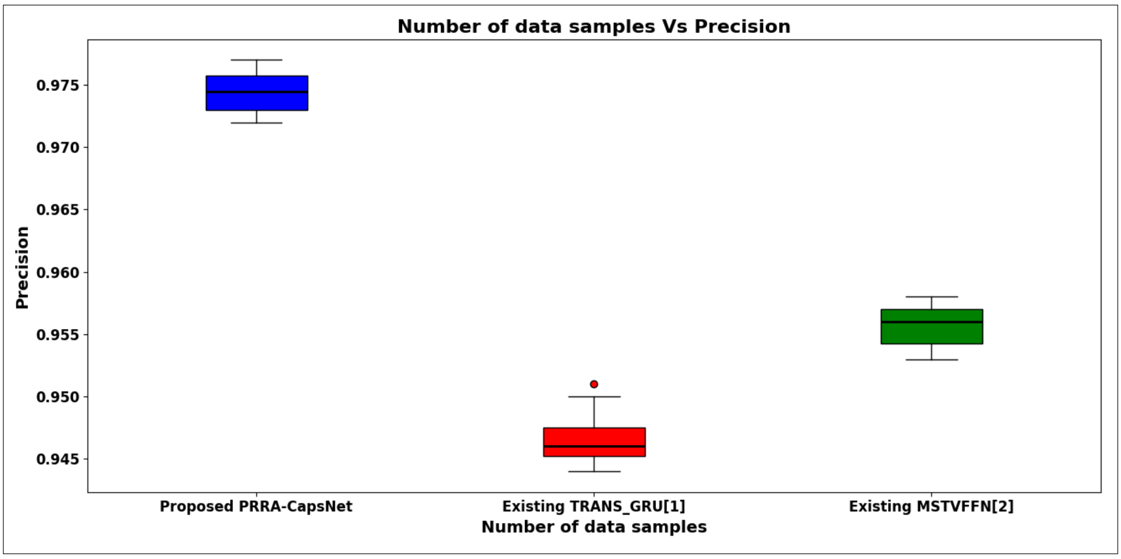

Figure 10 Graphical representation of precision

Figure 10 demonstrates the graphical illustration of precision respecting the number of data samples taken from the dataset ranges from 2000 to 20000. The overall presentation of precision performance in air pollution prediction is measured using three methods namely PRRA-CapsNet model outperforms the existing approaches TRANS_GRU [1], MSTVFFN [2]. The x-axis represents the number data samples, whereas the y axis assigns the overall precision performance. Among three different methods, PRRA-CapsNet model provides the superior performance in air pollution prediction than the other two existing methods. Let us consider

2000 data samples to evaluate the precision. In order to apply the PRRA-CapsNet model achieved a precision of 0.976, while the existing methods [1] and [2] observed the precision performance of 0.951 and 0.957 respectively. Similarly, divergent precision outcomes were accomplished relating to various calculations of input data samples. In conclusion, the overall observed results of PRRA-CapsNet model are compared to existing methods [1] and [2]. The average of ten comparison results specifies that the PRRA-CapsNet model demonstrates an improvement in precision of approximately 3% compared to [1] and compared to 2% over method [2]. This enhanced performance of PRRA-CapsNet model is achieved by integrating the Deep Capsule Network in grouping with the Kulczynski similarity function for data analysis. In addition, fine-tuning of Deep Capsule Network is done through the integration of Momentum method and adaptive horse herd optimization algorithm to reduce the training and justification errors. This algorithm effectively increases the performance of precision in air pollution prediction by increasing the true positive rate while minimizing the false positives.

5.3 Performance comparison of Recall

It also referred to as sensitivity, it measures accuracy of the model identifies positive instances during the classification. The formula for calculating recall is shown below:

$$Recall = \frac{TP}{TP+FN} \quad (26)$$

Where, true positive rate (TP), false negative 'FN'

Table 4 recall

Number of data samples	recall		
	Proposed PRRA-CapsNet	**Existing TRANS_GRU [1]**	**Existing MSTVFFN [2]**
2000	0.985	0.957	0.963
4000	0.982	0.954	0.962
6000	0.983	0.952	0.963
8000	0.985	0.953	0.964
10000	0.984	0.954	0.965
12000	0.983	0.953	0.962
14000	0.981	0.952	0.963
16000	0.983	0.951	0.964
18000	0.982	0.949	0.963
20000	0.981	0.946	0.965

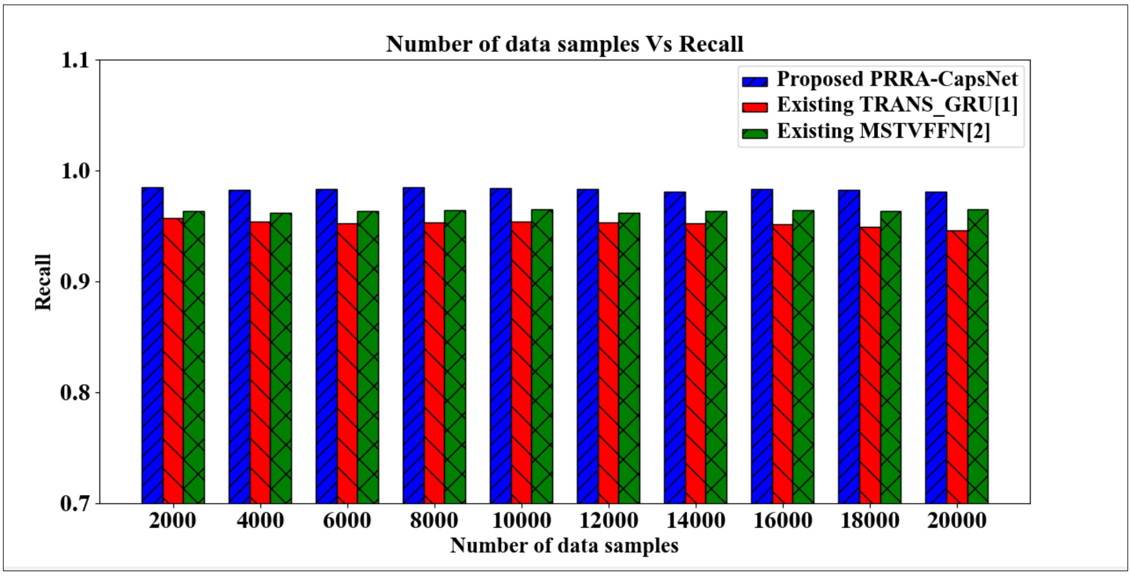

Figure 11 Graphical representation of recall

Figure 11 demonstrates an overall comparison analysis of recall performance with against the number of data samples collected from the air pollution dataset. The PRRA-CapsNet model reveals advanced recall performance in air pollution prediction compared to the existing approaches [1] and [2]. For instance, with 2000 data samples, PRRA-CapsNet model achieved a recall of 0.985, whereas methods [1] and [2] observed the recall rate of 0.957 and 0.963, respectively. Similarly, different performance outcomes were analyzed with dissimilar counts of input samples. The experimental assessment shows that the average recall performance of PRRA-CapsNet model improved by approximately 3% compared to [1] and 2% compared to [2]. The superior performance of PRRA-CapsNet model is attained owing to the fine-tuning process applied within the Deep Capsule Network. By utilizing this network architecture, the model decreases the error between predicted and actual air pollution prediction outcomes in multiclass output using the horse herd optimization algorithm. This process resulting in minimal false negatives and enhance in true positives for precise air pollution prediction.

5.4 Performance comparison of F1score

It also known F-measure calculated as the harmonic mean of both precision as well as recall. It is mathematically computed as follows

$$F1 - Score = 2 * \left(\frac{Precision * Recall}{Precision + Recall}\right) \quad (27)$$

Table 5 F1 score

Number of data samples	F1 score		
	Proposed PRRA-CapsNet	Existing TRANS_GRU [1]	Existing MSTVFFN [2]
2000	0.980	0.953	0.959
4000	0.978	0.951	0.958

6000	0.978	0.949	0.960
8000	0.978	0.949	0.960
10000	0.977	0.949	0.960
12000	0.979	0.949	0.959
14000	0.978	0.947	0.957
16000	0.978	0.948	0.958
18000	0.977	0.946	0.958
20000	0.976	0.946	0.958

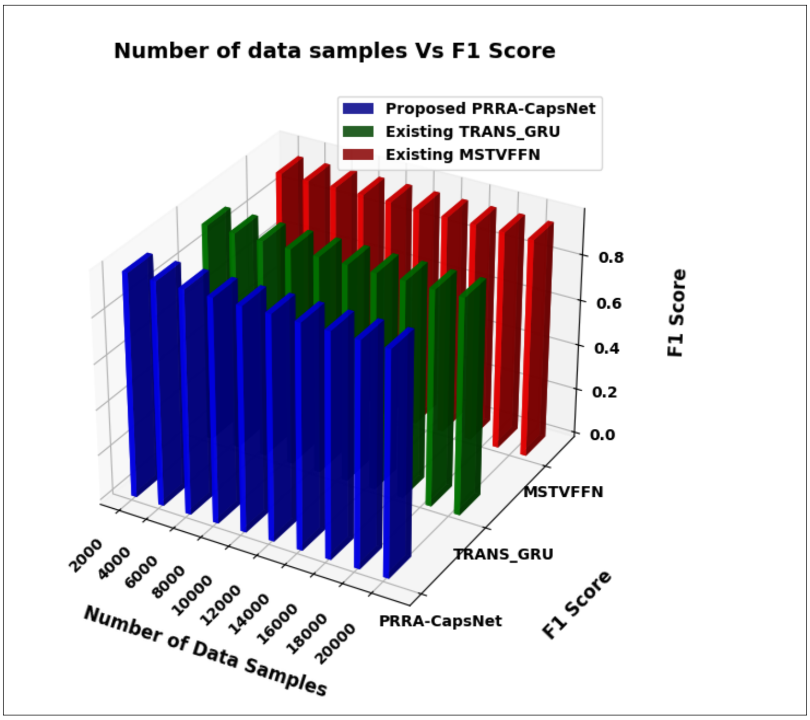

Figure 12 Graphical representations of F1 score

The graphical depiction of F1 scores performance analysis using three different methods namely PRRA-CapsNet model and existing approaches TRANS_GRU [1], MSTVFFN [2] are shown in figure 12. The result visibly shows that the higher F1 score is attained using PRRA-CapsNet model compared to other existing methods. In the first run with 2000 data samples, the F1 score of proposed PRRA-CapsNet model was observed to be 0.980, whereas the F1 score of existing methods [1] and [2] were 0.953 and 0.959 correspondingly. These outcomes confirm that the proposed PRRA-CapsNet model significantly enhances the F1 score compared to existing techniques. This development is achieved by the PRRA-CapsNet model increases the precision as well as recall in the air pollution prediction. Therefore, the performance of PRRA-CapsNet model was compared with the various results observed from the existing methods. Therefore, the overall results confirms that PRRA-CapsNet model increases the F1 score by 3% and 2% compared to [1] and [2] respectively.

5.5 Performance comparison of Root Mean square error

It is measured as the ratio of data samples in which the predicted results match the actual outcomes to the total number of data samples within the dataset. The RMSE is expressed as given below,

$$RMSE = \left[\sqrt{\frac{(Y_{actual} - Y_{predicted})^2}{n}} \right] \quad (28)$$

Where, $RMSE$ indicates a root mean square error, Y_{actual} represents the data samples for which actual results, $Y_{predicted}$ designates a predicted output.

Table 6 root mean square error

Number of data samples	RMSE		
	Proposed PRRA-CapsNet	Existing TRANS_GRU [1]	Existing MSTVFFN [2]
2000	0.072	0.167	0.145
4000	0.057	0.116	0.104
6000	0.044	0.101	0.085
8000	0.041	0.085	0.061
10000	0.035	0.075	0.057
12000	0.033	0.069	0.058
14000	0.032	0.064	0.048
16000	0.024	0.060	0.045
18000	0.026	0.056	0.042
20000	0.025	0.053	0.045

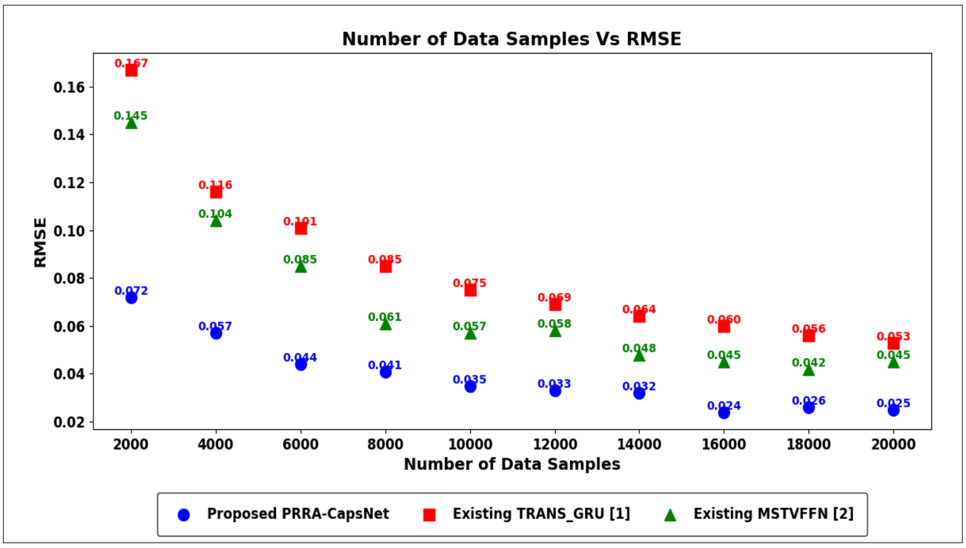

Figure 13 Graphical representations of RMSE

Figure 13 exhibits the various performance outcomes of RMSE for air pollution prediction across dissimilar sample sizes, ranging from 2000 to 20000 data samples. As shown in the figure 13, the PRRA-CapsNet model consistently achieved smaller RMSE when compared to other existing deep learning architectures. However, the average of ten outcomes indicates that the PRRA-CapsNet model demonstrates a 54% reduction in RMSE compared to

[1] and a 42% reduction relative to [2]. This well-known improvement is achieved due to incorporation of the horse herd optimization algorithm within the deep capsule network. The optimization algorithm efficiently fine-tunes the model hyperparameters, thereby lessening the air pollution prediction errors. This process integrates the deep capsule network and optimization techniques enable PRRA-CapsNet to notably reduce the error rate in air pollution prediction.

5.6 Performance comparison of mean absolute error

It is accuracy parameter used to compute the amount of the errors to the average value of the predicted results. The mean absolute error is mathematically computed as follows,

$$MAE = \sum \left| Y_{actual} - Y_{predicted} \right| \quad (29)$$

Where, MAE denotes mean absolute error, Y_{Actual} denotes the actual results of weather forecasting, $Y_{predicted}$ denotes a predicted results.

Table 7 means absolute error

Number of data samples	MAE		
	Proposed PRRA-CapsNet	Existing TRANS_GRU [1]	Existing MSTVFFN [2]
2000	3.25	7.5	6.5
4000	3.63	7.35	6.58
6000	3.45	7.85	6.64
8000	3.75	7.65	5.48
10000	3.56	7.55	5.78
12000	3.72	7.65	6.37
14000	3.83	7.58	5.79
16000	3.15	7.65	5.78
18000	3.58	7.6	5.7
20000	3.67	7.53	6.37

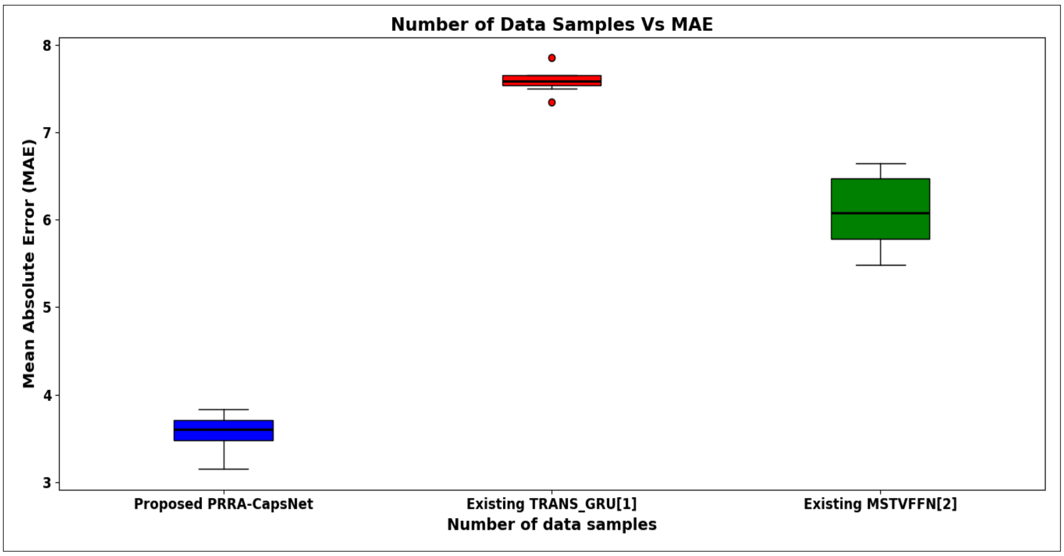

Figure 14 Graphical representations of MAE

Figure 14 given above reveals a graphical assessment of the MAE by means of varying data sample sizes ranging from 2000 to 20000. The evaluation involved three various deep learning models namely the proposed PRRA-CapsNet model and existing approaches TRANS_GRU [1], MSTVFFN [2]. The x-axis indicates the number of data samples, while y-axis displays the equivalent MAE values. The results obviously demonstrate that the PRRA-CapsNet model consistently delivers the smaller MAE compared to the other existing deep learning methods. For instance, let us considered as 2000 data samples in first iteration, the performance of MAE achieved by PRRA-CapsNet model is 3.25, whereas [1] [2] observed the MAE to be 7.5 and 6.5 respectively. The overall results underscore that PRRA-CapsNet model declines MAE by approximately 53% compared to [1] and by 41% relative to [2], enhancing air pollution prediction accuracy. This development is primarily due to the employment of the horse herd optimization algorithm. The fine-tuning process of deep capsule network efficiently decreases the prediction errors, thereby lessening the overall MAE.

5.7 Performance comparison of prediction time

It refers to the amount of time consumed by the algorithm to perform the prediction from the given input data samples. Therefore, total prediction time is mathematically calculated as follows.

$$PRT = \sum_{i=1}^{n} SR_i * TM\ [AP] \quad (30)$$

Where, PRT designates prediction time, $TM\ [AP]$ indicates time for prediction with respect to number of data samples 'SR_i'. The overall prediction time is measured in terms of milliseconds (ms).

Table 8 prediction time

Number of data samples	prediction time (ms)		
	Proposed PRRA-CapsNet	Existing TRANS_GRU [1]	Existing MSTVFFN [2]
2000	54	68	65
4000	55.8	73.6	67.5
6000	60.2	78.6	71.4
8000	63.5	84.5	75.8
10000	65.8	90.6	80.2
12000	70.3	96.8	88.3
14000	72.6	103.6	93.5
16000	76.8	109.7	98.4
18000	82.5	112.2	102.4
20000	88.4	116.8	108

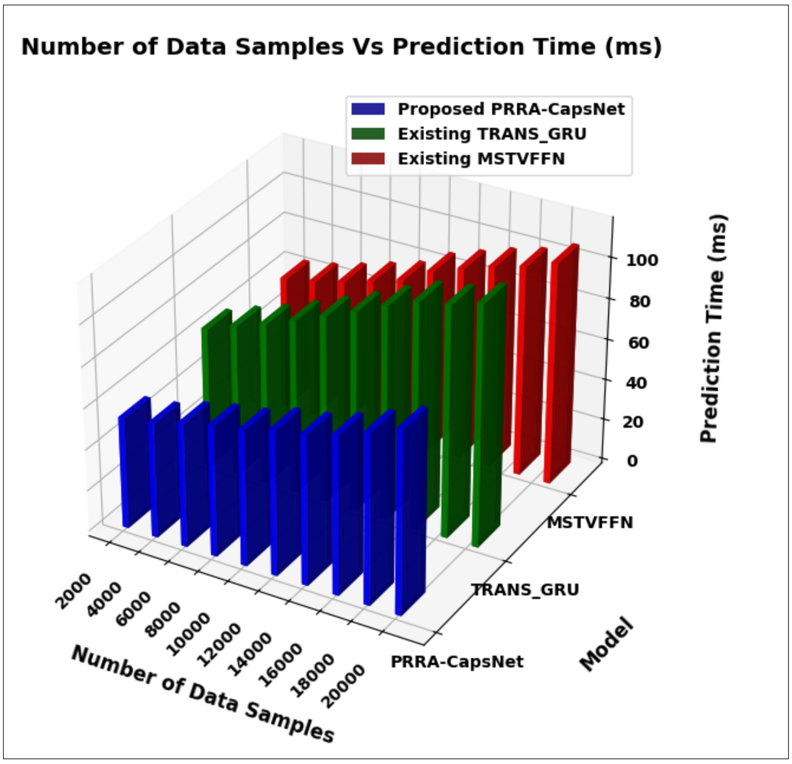

Figure 15 Graphical analysis of prediction time

Figure 15 demonstrates the performance outcomes of air pollution prediction time using the proposed PRRA-CapsNet model and two existing approaches, TRANS_GRU [1], MSTVFFN [2]. The comprehensive graphical charts are observed against the number of input data samples, ranging from 2000 to 20000. Since the number of data sample increases, the air pollution prediction for all three methods increases consequently. The horizontal axis indicates the number of data samples gathered from the dataset, while the vertical axis indicates the corresponding air pollution prediction time. In an experiment with 2000 data samples, PRRA-CapsNet model achieved a air pollution prediction time of $54ms$, while methods [1] and [2] observed a prediction time of $68ms$ and $65ms$ respectively. Among three various methods, the PRRA-CapsNet model constantly reduces the air pollution prediction time by 26% and 19% when compared to the two existing deep learning models. This effectiveness is achieved due to its data pre-processing and relevant feature selection process. In the data preprocessing, Polynomial predictive mean imputation method is employed for handling the missing data. Followed by, Modified Z-Score method is used for outlier data removal. Moreover, the robust autoencoder based feature selection process is carried out in PRRA-CapsNet model continually to choose the more relevant features for accurate air pollution prediction. Overall, the analysis confirms that the PRRA-CapsNet model reduced prediction time in air pollution prediction.

6. Conclusion

This paper presents a PRRA-CapsNet model to enhance the accuracy of predictions in cloud computing environment. In this paper, a novel PRRA-CapsNet model specifically developed to provide highly accurate predictions with minimal time as well as error rate. In PRRA-CapsNet model, the **d**eep **c**apsule network introduces the data preprocessing and the selection of important features from the dataset, aiming to decreasing the overall time required for prediction. This method utilizes the model's capability to deep **c**apsule network, enabling proficient analysis of data samples using Kulczynski similarity function to enhance the accuracy of prediction. Furthermore, error rate in the predictive analytics process is further minimized through fine-tuning process with horse herd optimization algorithm. An inclusive statistical assessment is carried out by means of numerous performance metrics, such as prediction accuracy, precision, recall, F1-score, RMSE, MAE, and prediction time. The experimental outcomes reveal that the PRRA-CapsNet model consistently outperforms predictable deep learning approaches by delivering higher prediction accuracy, faster times, and minimized the error rates.

References

[1] Sudhir Kumar, Vaneet Kour, Ankit Raj, Tagru Tapung, Shivendu Mishra, Rajiv Misra, "Optimizing Air Pollution Forecasting Models Through Knowledge Distillation: A Novel GCN and TRANS_GRU Methodology for Indian Cities", IEEE Access , Volume 13, 2025, Pages 40237 – 40257. **DOI:** 10.1109/ACCESS.2025.3546504

[2] Xinmeng Zhou , Xun Liang , Qiqi Zhu , Junfeng Gu, Qingfeng Guan, "A multiscale spatial–temporal-variable feature fusion network for predicting multiple air pollutants", Environment International, Elsevier, Volume 205, November 2025, Pages 1-16. https://doi.org/10.1016/j.envint.2025.109864

[3] Gang Chen, Shen Chen, Dong Li & Cai Chen, "A hybrid deep learning air pollution prediction approach based on neighborhood selection and spatio-temporal attention", Scientific Reports, volume 15, 2025, Pages 1-20. https://doi.org/10.1038/s41598-025-88086-1

[4] Sabina-Cristiana Necula, Ileana Hauer, Doina Fotache and Luminit, Hurbean, "Advanced Hybrid Models for Air Pollution Forecasting: Combining SARIMA and BiLSTM Architectures", Electronics, Volume 14, Issue 3, 2025, Pages 1-38. https://doi.org/10.3390/electronics14030549

[5] Cog Cao, Ramit Debnath and R Michael Alvarez, "Physics-based machine learning for predicting urban air pollution using decadal time series data", Environmental Research Communications, Volume 7, 2025, Pages 1-9. **DOI** 10.1088/2515-7620/add795

[6] Cheng-Hong Yang, Po-Hung Chen, Cheng-San Yang, Li-Yeh Chuang, "Analysis and Forecasting of Air Pollution on Nitrogen Dioxide and Sulfur Dioxide Using Deep Learning", IEEE Access, Volume 12, 2024, Pages 165236 – 165252. **DOI:** 10.1109/ACCESS.2024.3494263

[7]Pranshu Patel, Swara Patel , Kanish Shah , Kedar Desai, Samrat Patel, Manan Shah, Samir Patel, " A systematic study on PM2.5 and PM10 concentration prediction in air pollution using machine learning and deep learning model", Environmental Chemistry and Ecotoxicology, Elsevier, Volume 7, 2025, Pages 1401-1415. https://doi.org/10.1016/j.enceco.2025.07.001

[8] Mostafa M. Abdelmalek, Hatem Mahmoud & Hassan Shokry, "Prognosis of air quality index and air pollution using machine learning techniques" Scientific Reports, volume 15, 2025, Pages 1-23. https://doi.org/10.1038/s41598-025-11260-y

[9] Jiankun Zhang, Pei Su, JuexuanWang and Zhantong Cai, "Air Pollutant Concentration Prediction Using a Generative Adversarial Network with Multi-Scale Convolutional Long Short-Term Memory and Enhanced U-Net", Sustainability, Volume 17, Issue 24, 2025, Pages 1-22. https://doi.org/10.3390/su172411177

[10] Miguel Angel Solis Teran, Felipe Leite Coelho da Silva, Elias A. Torres Armas, Natali Carbo-Bustinza and Javier Linkolk Lopez-Gonzales, "Modeling Air Pollution in Metropolitan Lima: A Statistical and Artificial Neural Network Approach", Environments, Volume 12, Issue 6, 2025, Pages 1-13. https://doi.org/10.3390/environments12060196
[11] Xinyi Mao, Gen Liu, JianWang and Yongbo Lai, "BiTCN-ISInformer: A Parallel Model for Regional Air Pollutant Concentration Prediction Using Bidirectional Temporal Convolutional Network and Enhanced Informer", Sustainability, Volume 17, Issue 19, 2025, Pages 1-27. https://doi.org/10.3390/su17198631
[12] Jie Liu, Fanshu Ding, Xiuyi Yao, Wanqing Wang, Ziyu Chen, Xiaolin Yang, Haoyuan Zhu, "Decomposition and forecasting of concentrations of highly volatile atmospheric pollutants in plateau cities based on the VMD-CNN-BiLSTM framework", Environmental Technology & Innovation, Elsevier, Volume 41, March 2026, Pages 1-24. https://doi.org/10.1016/j.eti.2025.104717
[13] Fareena Naz, Muhammad Fahim, Adnan Ahmad Cheema, Bradley D. E. Mcniven ,Tuan-Vu Cao; Ruth Hunter, "Air Quality and Healthy Ageing: Predictive Modeling of Pollutants Using CNN Quantum-LSTM", IEEE Access , Volume 13, 2025, Pages 94212 – 94223. **DOI:** 10.1109/ACCESS.2025.3570526
[14] Yurii Andrashko, Oleksandr Kuchanskyi, Andrii Biloshchytskyi, Alexandr Neftissov and Svitlana Biloshchytska, "Forecasting Air Pollutant Emissions Using Deep Sparse Transformer Networks: A Case Study of the Ekibastuz Coal-Fired Power Plant", Sustainability, Volume 17, Issue 11, 2025, Pages 1-17. https://doi.org/10.3390/su17115115
[15] Rafat Zrieq, Souad Kamel, Faris Al-Hamazani, Sahbi Boubaker, Rozan Attili, and Marcos J. Arauzo-Bravo, "Toxics, Volume 13, Issue 8, 2025, Pages 1-31. https://doi.org/10.3390/toxics13080682
[16] Yunzhu Liu, "Deep Aggregation seq2seq Network With Time Feature Fusion for Air Pollutant Concentration Prediction in Smart Cities", Engineering Reports, Wiley, Volume 7, Issue 2, 2025, Pages 1-14. https://doi.org/10.1002/eng2.70031
[17] Rui Zhang & Norhashidah Awang, "A novel ST-iTransformer model for spatio-temporal ambient air pollution forecasting", Journal of Big Data, Springer, Volume 12, 2025, Pages 1-27. https://doi.org/10.1186/s40537-025-01150-5
[18] Huazhen Xu,Wei Song,Lanmei Qian,Xiangxiang Mei,Guojian Zou, "STGATN: A novel spatiotemporal graph attention network for predicting pollutant concentrations at multiple stations", PLoS One, Volume 20, Issue 7, 2025, Pages 1-26. https://doi.org/10.1371/journal.pone.0328532
[19] Vibha Yadav, Amit Kumar Yadav & Rohit Khargotra, "Relief-Based Feature Selection for ANN-Driven Air Pollution Prediction and School Safety Policy in India", Water, Air, & Soil Pollution, Springer, Volume 237, 2026, Pages 1-18. https://doi.org/10.1007/s11270-025-08759-5
[20] Abdel Salam Alsabagh, Omer A. Alawi, Haslinda Mohamed Kamar, Ahmed Adil Nafea, Mohammed M. AL-Ani, Hussein A. Mohammed, S. N. Kazi, Atheer Y. Oudah & Zaher Mundher Yaseen, "Deep learning framework for hourly air pollutants forecasting using encoding cyclical features across multiple monitoring sites in Beijing", Scientific Reports, volume 15, 2025, Pages 1-33. https://doi.org/10.1038/s41598-025-05472-5

[21] Sunny Kumar, Arpita Vishwakarma, Manish Kumar Srivastava, Dr. Yusuf Perwej, Dr. Nikhat Akhtar, "Ensemble Machine Learning for Reliable Air Pollution Prediction and Sustainable Environmental Management", International Journal of Scientific Research in Science and Technology, Volume 13, Issue 1, 2026, Pages 53-67. https://doi.org/10.32628/IJSRST261316

[22] Nor Irwin Basir,Kathlyn Kaiyun Tan,Danny Hartanto Djarum,Zainal Ahmad,Dai-Viet N.Vo, Jie Zhang, "Autoencoder Artificialneuralnetwork Model For Air Pollution Index Prediction", IIUM Engineering Journal, Volume 26, Issue 1, 2025, https://doi.org/10.31436/iiumej.v26i1.2818

[23] Iman Zandi, Ali Jafari and Aynaz Lotfata, "Enhancing PM2.5 Air Pollution Prediction Performance by Optimizing the Echo State Network (ESN) Deep Learning Model Using New Metaheuristic Algorithms", Urban Science, Volume 9, Issue 5, 2025, Pages 1-19. https://doi.org/10.3390/urbansci9050138

[24] Musa Milli, "Designing a residual-enhanced hybrid Prophet–LSTM framework for urban air pollution forecasting in Beijing", Scientific Reports, volume 15, 2025, Pages 1-24. https://doi.org/10.1038/s41598-025-27510-y

[25] Alejandra Abalo-García, Sergio Hernández-García, Iván Ramírez, Emanuele Schiavi, "MPD: A Meteorological and Pollution Dataset: A Comprehensive Study of Machine and Deep Learning Methods for Air Pollution Forecasting", IEEE Access, Volume 13, 2025, Pages 41282 – 41299. **DOI:** 10.1109/ACCESS.2025.3547038

[26] Alireza Zhalehdoost & Mohammad Taleai, "Unravelling the importance of spatial and temporal resolutions in modeling urban air pollution using a machine learning approach", Scientific Reports, volume 15, 2025, Pages 1-26. https://doi.org/10.1038/s41598-025-13639-3

[27] Youness El Mghouchi and Mihaela Tinca Udristioiu , "Integrated AI Framework for Sustainable Environmental Management: Multivariate Air Pollution Interpretation and Prediction Using Ensemble and Deep Learning Models", Sustainability, Volume 18, Issue 3, 2026, Pages 1-28. https://doi.org/10.3390/su18031457

[28] Nadia AL-Rousan, Hazem Al-Najjar, Ismail A. Elhaty, "Machine learning framework for forecasting air pollution: Evaluating seasonal and climatic influences in Istanbul, Turkey", PLoS One, Volume 20, Issue 10, 2025, Pages 1-22. https://doi.org/10.1371/journal.pone.0330716

[29] Matteo Sangiorgio, Giorgio Guariso, "Local vs regional neural air pollution forecasting models", IFAC Journal of Systems and Control, Elsevier, Volume 31, March 2025, Pages 1-10. https://doi.org/10.1016/j.ifacsc.2025.100298

[30] Nurtiti Sunusi, Ankaz As Sikiba , Sumanta Pasari, "A novel hybrid CLARA and fuzzy time series Markov chain model for predicting air pollution in Jakarta", MethodsX, Elsevier, Volume 14, June 2025, Pages 1-11. https://doi.org/10.1016/j.mex.2025.103202

Authors Profile

M.Nandhiya received her B.Sc Information Technology from NGM College, Pollachi. She had her Master of Computer Applications, from SVS Insitute of Technology under Anna University, Chennai. She holds her M.Phil degree in Computer Science at NGM College, Pollachi. She had 2 years 7 months experience in the field of teaching. She is presently working as a Assistant professor in NGM College, Pollachi. Her research interest includes Data Mining, Big Data Analytics. Now she is pursing her PhD degree at NGM College, Pollachi.

Dr. C.R. Durga devi, Assistant Professor in the Department of Information Technology at Nallamuthu Gounder Mahalingam College, Pollachi, is an accomplished academic with over 18 years of teaching and research experience. She holds a Ph.D. in Computer Science (2022) from Bharathiar University and has qualified both NET (2013) and SET (2012). Her research interests include cloud computing, load balancing, big data analytics, and optimization algorithms, with 4 Scopus indexed publications, 10international journal papers, and 20 conference presentations. Dedicated to continuous learning, she has completed 25 FDPs, 12 workshops, and industry trainings with Wipro on Java Full Stack and Cloud Computing. A recognized Ph.D. supervisor, she contributes as a doctoral committee member and evaluator while also serving as Criterion Manager for NAAC, member of the Institution's Innovation Council, and organizer of innovation-focused initiatives. She has been recognized as a Wipro Certified Faculty (2022)and Innovation Ambassador (MIC, 2024–25), and actively engages in academic responsibilities such as external examiner, question paper setter, curriculum designer, and editor. Frequently invited as a resource person, she exemplifies academic excellence, research leadership, and a strong commitment to advancing higher education in Information Technology.