IoT Enabled Robust Deep Capsule Network for Cloud-Based Predictive Analytics

Mrs. M. Nandhiya[1], Dr. C. R Durga Devi[2]
[1]*Ph.D Research Scholar, Department of Computer Science, NGM College, Pollachi, Tamil Nadu, India*
[2]*Assistant Professor, Department of Information Technology, NGM College, Pollachi, Tamil Nadu, India*

***Abstract*—Cloud computing provides on-demand access to scalable resources, including storage, networking, analytics, and software applications through internet-based platforms. Predictive analytics has become a significant cloud computing application, enabling the estimation of future outcomes using historical and real-time data. However, conventional prediction techniques often experience reduced accuracy and increased computational overhead when processing large-scale and heterogeneous datasets.**

This study proposes an IoT-enabled Robust Autoencoded Deep Capsule Network to enhance predictive analytics in cloud computing environments. The developed framework combines data preprocessing, feature extraction, classification, and optimization within an integrated architecture. IoT devices continuously collect environmental information and transmit it to cloud servers for further processing and analysis. During the preprocessing stage, missing and anomalous data are addressed using polynomial-based imputation and modified Z-score techniques. Subsequently, a robust autoencoder extracts significant features, thereby reducing data dimensionality and enhancing computational efficiency. The extracted features are then processed by a capsule network that employs a similarity-driven classification mechanism to produce reliable prediction outcomes.

Performance assessment conducted on the Global Air Pollution Dataset indicates that the proposed approach outperforms existing deep learning models in terms of prediction accuracy while simultaneously reducing prediction errors and execution time.

***Index Terms*—IoT, Cloud Computing, Predictive Analytics, Deep Capsule Network, Autoencoder, Feature Selection, Air Pollution Prediction, Machine Learning**

I. INTRODUCTION

Cloud computing has emerged as a vital technology for providing scalable computing resources and services through the internet. Many organizations utilize cloud platforms to store, manage, and process vast volumes of data without investing in dedicated physical infrastructure. The convergence of Internet of Things (IoT) devices with cloud environments has facilitated real-time data collection, monitoring, and intelligent decision-making across domains such as healthcare, agriculture, environmental surveillance, and smart city applications.

In recent years, machine learning and deep learning approaches have been widely adopted for predictive analytics tasks. Techniques including LSTM, BiLSTM, CNN, Transformer, and various hybrid neural network models have demonstrated significant improvements in forecasting accuracy. Despite these advancements, several existing approaches demand substantial training time and encounter difficulties when handling high-dimensional data. Furthermore, some models are unable to effectively learn both spatial and temporal dependencies present within complex datasets.

To overcome these challenges, this study introduces a Polynomial Regressive Robust Autoencoded Deep Capsule Network (PRRA-CapsNet). The proposed framework is designed to enhance prediction performance while reducing computational overhead. By integrating data preprocessing, feature selection, classification, and optimization mechanisms within a unified architecture, PRRA-CapsNet provides an efficient solution for cloud-based predictive analytics

II. RELATED WORKS

Several researchers have developed advanced approaches for air pollution prediction and predictive analytics. Transformer-GRU models were introduced to improve forecasting performance, but they faced scalability issues with large datasets. Spatial-temporal feature fusion networks improved pollutant prediction but required extensive parameter tuning. Deep learning frameworks based on CNN, LSTM, GAN, and attention mechanisms have also been explored for environmental monitoring.

Autoencoder-based methods were proposed to reduce dimensionality and enhance feature extraction. Although these techniques improved learning efficiency, they often increased training complexity. Hybrid forecasting models combined machine learning and statistical methods to improve prediction accuracy, but many failed to reduce prediction time and error rates simultaneously.

The review of earlier studies shows that most existing models either suffer from high computational cost or reduced generalization capability. Therefore, a lightweight and accurate prediction framework is necessary for cloud environments handling large-scale IoT data.

III. PROPOSED METHODOLOGY

The PRRA-CapsNet framework is designed to perform efficient predictive analytics in cloud computing environments. The architecture consists of four major stages: data acquisition, preprocessing, feature selection, and classification.

Initially, IoT sensors collect environmental information from multiple locations. These data samples are transmitted to the cloud server for storage and analysis. The collected dataset is then passed to the preprocessing stage. During preprocessing, missing values are estimated using polynomial predictive mean imputation. Outlier samples are detected and removed through the modified Z-score technique. This process improves the quality of the dataset and reduces noise.

After preprocessing, a robust auto encoder performs feature selection. The encoder transforms the high-dimensional input into a compact latent representation, while the decoder reconstructs the original data. The auto encoder identifies relevant features and removes redundant attributes. This dimensionality reduction process decreases computational overhead and enhances prediction performance.

The selected features are then provided to the class capsule layer. Capsule networks capture spatial relationships between features more effectively than traditional neural networks.

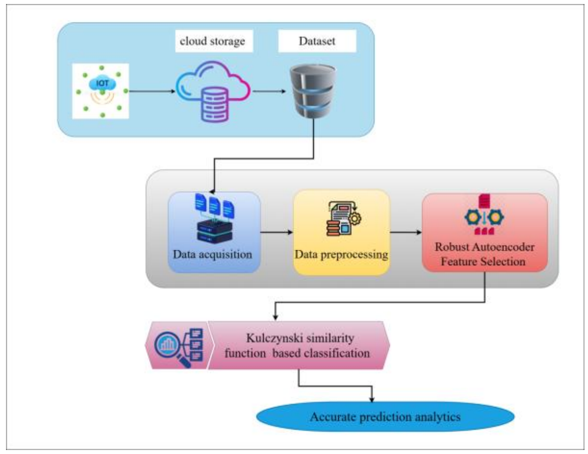

Figure 1. Architecture of the proposed PRRA-CapsNet model

The proposed model employs the Kulczynski similarity function to measure similarity between training and testing samples for accurate classification. To minimize classification error, the network parameters are optimized using a wild horse herd optimization algorithm. The optimization process iteratively updates weight values to achieve better prediction accuracy and lower false prediction rates.

IV. EXPERIMENTAL SETUP

The proposed PRRA-CapsNet model was implemented using Python programming language. Experiments were conducted on the Global Air Pollution Dataset obtained from the Kaggle repository. The dataset contains 23,463 instances with 12 attributes related to air quality, including AQI value, ozone level, nitrogen dioxide concentration, carbon monoxide, and particulate matter.

The preprocessing stage reduced dataset noise and removed abnormal values. Feature selection further identified the most significant attributes for prediction. Seven major features were selected for final analysis.

The model performance was evaluated using metrics such as prediction accuracy, precision, recall, F1-score, RMSE, MAE, and prediction time. Existing methods including TRANS_GRU and MSTVFFN were used for comparison.

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 2 data acquisition

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 3 after data preprocessing

V. RESULTS AND DISCUSSION

Experimental results indicate that the proposed PRRA-CapsNet framework outperforms existing deep learning methods in predictive analytics tasks. The model achieved higher prediction accuracy because of efficient preprocessing and similarity-based classification.

The robust auto encoder improved feature extraction by removing irrelevant attributes and reducing dimensionality. As a result, computational complexity and prediction time were significantly reduced. The capsule network effectively captured feature relationships, leading to improved classification accuracy.

The optimization strategy based on wild horse herd behavior minimized prediction error and improved convergence. Compared with TRANS_GRU and MSTVFFN models, the proposed approach demonstrated better precision, recall, and F1-score. In addition, the framework achieved lower RMSE and MAE values, confirming its reliability for cloud-based predictive analytics applications.

VI. CONCLUSION

This paper introduced a robust IoT-enabled predictive analytics framework called PRRA-CapsNet for cloud computing environments. The model integrates preprocessing, robust autoencoder feature selection, deep capsule classification, and optimization into a unified architecture. By combining these techniques, the framework improves prediction accuracy while reducing computational complexity and prediction time.

Experimental analysis using the Global Air Pollution Dataset confirmed that the proposed model provides superior performance compared with existing deep learning approaches. Future work can focus on integrating real-time big data analytics and advanced optimization strategies for smart city and environmental monitoring applications.

REFERENCES

[1] Sudhir Kumar *et al.*, "Optimizing Air Pollution Forecasting Models Through Knowledge Distillation," *IEEE Access*, 2025.

[2] Xinmeng Zhou et al., "A Multiscale Spatial-Temporal Variable Feature Fusion Network," Environment International, 2025.

[3] Gang Chen et al., "Hybrid Deep Learning Air Pollution Prediction," Scientific Reports, 2025.

[4] Sabina-Cristiana Necula et al., "Advanced Hybrid Models for Air Pollution Forecasting," Electronics, 2025.

[5] Cog Cao *et al.*, "Physics-Based Machine Learning for Urban Air Pollution Prediction," *Environmental Research Communications*, 2025.

AUTHORS PROFILE

M. Nandhiya received her B.Sc. Information Technology from NGM College, Pollachi. She had her Master of Computer Applications, from SVS Institute 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 an Assistant professor in NGM College, Pollachi. Her research interest includes Data Mining, Big Data Analytics. Now she is pursuing 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, 10 international 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.