

RESEARCH ARTICLE

DEEP LEARNING-DRIVEN AIR QUALITY MONITORING USING VAE AND GNN MODEL FOR REAL-TIME MOBILE ACCESSIBILITY

Raju Barskar ^{1,*} and A. Vijayaraj ²

¹ Associate Professor, Department of Computer Science & Engineering, University Institute of Technology, RGPV Bhopal (M.P.), Madhya Pradesh, India.

² Associate Professor, Department of Information Technology, R.M.K Engineering College, RSM Nagar, Kavaraipettai, Chennai, Tamil Nadu, India.

*Corresponding e-mail: rajubaraskar@rgpv.ac.in

Abstract - Monitoring air quality has become one of the most essential activities in IoT industrial and urban areas. The quality of air is adversely affected due to various forms of pollution caused by transportation, electricity, fuel uses. Air quality monitoring enables early detection of harmful pollutants, allowing for timely interventions to protect public health and improves the quality of life. However, traditional way of using fixed sensors cannot effectively provide a comprehensive view of air pollution in people's surroundings. To overcome these issues, a novel GNN Model Based Air quality monitoring (GNN-MBA) has been proposed to enable real-time prediction of air quality and displays the quality results to user's mobile devices. The proposed method utilizing the Variational Autoencoder (VAE) for feature extraction to efficiently capture features for accurate classification. The extracted features are fed into Graph Neural Network (GNN) deep learning model, which categorizes the data into three classes such as Good, Moderate and Poor, the result is then forwards to gateway. The gateway sends the processed output to WiFi network, making the air quality information accessible to the user's mobile device. Measures including F1score (F1S), accuracy, precision and recall are utilized to assess the suggested approach. The GNN-MBA accuracy in the Beijing Air Quality dataset is 0.7%, 2.5%, and 3.2% greater than the current ICEEMDAN-WOA-ELM and SARIMA approaches, while its RMSE is decreased by 4.5%, 1.5%, and 2.0% respectively.

Keywords – Air quality monitoring, Variational Autoencoder, Graph Neural Network, Deep learning.

1. INTRODUCTION

ISSN: 2584-1041

The Internet of Things (IoT) is a network of physical and virtual items that are linked together over the Internet, which has grown significantly in recent years thanks to breakthroughs in various areas, such as hardware, software, and networking technologies. With the proper deployment of IoT devices, sensors can accurately describe their surroundings, convey data, and communicate with one another. It is possible to understand the environment around us and respond to any scenario requiring immediate attention. Home monitoring and security systems, indoor and

outdoor air quality monitoring and management, energy management in public and private buildings, healthcare systems, and micro-climate monitoring systems in specific locations are some examples that IoT can help.

Air Pollution (AP) is one of the serious and major environmental problem worldwide. Many researchers have drawn attention and have focused about these problems keeping in mind human health. Air quality prediction information is one of the better ways through which people can be informed to be more vigilant about serious health issues and protect human health caused by air pollution. In many metropolitan cities air pollution is a major challenging environmental issue. To analyze the present traffic condition of the city, local authorities can be enabled by real time monitoring of pollution data which makes appropriate decisions. Hence an early system is required for monitoring and calculating the level of AP using Air Quality (AQ) which essential for predicting exactly the pollutant concentrations.

The air quality of a particular area can be monitored based devices and sensors Arduino/Raspberry Pi. Air Pollution is increasing heavily these days due to the many important factors like Vehicle Emissions, Deforestation, Industrialization. Old vehicles could produce more smoke and hence those vehicles could be banned from using. The purpose of this research study is to understand Information on environmental variables and also allowing easy integration into any other type of internetbased architecture (IoT) which allows the use of sensors capable of collect information on sensors related to smart city environment measurements, with a view to providing data on which environmental pollution-related information. The suggested work's primary contributions are as follows:

• The objective of the GNN-MBA approach is to provide an efficient air quality monitoring using DL to real-time monitoring, which is then used to get user awareness.

- To improve data quality, the collected data from sensors are first pre-processed using data cleaning, data normalization and data reduction.
- The GNN-MBA technique leverages feature extraction by using VAE to capture features efficiently and raise the classification model's accuracy.
- The extracted features are then used GNN deep learning model for classification, allowing it to correctly classify data into three classes such as Good, Moderate, Poor, the result is then sent to the gateway and forwarded via a Wi-Fi network to the user's mobile device.
- The effectiveness of suggested GNN-MBA technique is evaluated utilizing parameters like accuracy, recall (RC), f1-score, precision.

The document's remaining sections are organized as follows. The literature review is covered in Part II. The created air quality monitoring using DL is described detail in Part III. The results and observations from the experiment are shown in Part IV. The recommendations and conclusion are in Part V.

2. LITERATURE REVIEW

In 2022, Marzouk, M. and Atef, M., [14] suggested a system for monitoring different air parameters to evaluate the indoor air quality (IAQ) and to provide real-time readings. The average readings for temperature, humidity, air pressure, CO₂, CO, and PM_{2.5} in the presented case study are 30 °C, 42%, 100,422 pa, 460 ppm, 2.2 ppm, and 15.3 μ/m^3 , respectively.

In 2022, Tran, Q.A., et al [15] suggested an IoT-based Air Quality Monitoring and Forecasting System to monitor and predict air pollution for a specific area based on various pollution factors. The result shows that our system can predict the air quality factors over the next hour with the highest accuracy at 96%.

In 2023, Wu, C.L.et al [16] suggested a novel deep learning-based hybrid model of Res-GCN-BiLSTM combining the residual neural network (ResNet), graph convolutional network (GCN), and bidirectional long short-term memory (BiLSTM), for predicting short-term regional NO $_2$ and O $_3$ concentrations. This model improved prediction accuracy with a 11% and 17% reduction in MAE for NO $_2$ and O $_3$ compared to the best baseline.

In 2023, Wei, Y., et al [17] suggested a hybrid deep learning model that combines LSTM with Autoencoder for anomaly detection tasks in indoor air quality (IAQ) to address anomaly detection in the IAQ area. The result demonstrates a very high and robust accuracy rate (99.50%) that outperforms other similar models.

In 2023, Fu, L., Li, J. and Chen, Y., [18] suggested an improved complete ensemble empirical mode decomposition with adaptive noise-whale optimization algorithm-extreme learning machine (ICEEMDAN-WOA-ELM) to resolve real air quality monitoring challenges in environment. The

proposed method is 0.946, which are higher than those of the other models.

In 2023, Karnati, H., 2023. [19] suggested a development of a portable air quality detection device that can be used anywhere to address air Pollution Monitoring systems are used to measure the concentration of gases like CO2, smoke, alcohol, benzene, and NH3 present in the air.

In 2024, Ansari, M. and Alam, M., [20] suggested a novel BO-HyTS approach that combines seasonal autoregressive integrated moving average (SARIMA) and long short-term memory (LSTM) using Bayesian optimization to predict air pollution levels. Results show with an MSE of 632.200, RMSE of 25.14, Med AE of 19.11, Max Error of 51.52, and MAE of 20.49.

The evaluation of the literature indicates that current assistive technologies for air quality monitoring exhibit several key drawbacks such as limited mobile accessibility, reducing the accuracy and timeliness of predictions. This study presents a novel GNN-MBA, which will be covered in more detail in the next section, in order to address these issues

3. PROPOSED METHOD

In this section, a novel GNN-MBA technique has been proposed to monitor air quality accurately to improves the quality of life. Initially the datas are collected from air quality monitoring station using sensors and pre-processed utilizing methods such as Data cleaning, Data Normalization, Data reduction to enhance the data quality. After pre-processing the essential features are extracted using VAE model. Finaly the extracted features are fed into GNN model which classifies data into three classes such as Good, Moderate and Poor, the result is then sent to the gateway and forwarded via a Wi-Fi network to the user's mobile device to get early warnings. Figure 1 shows the proposed system's workflow.

3.1 Data Collection

Data collection involves measuring pollutants and environmental parameters with a variety of sensors and instruments. There is a combination of low-cost sensor technologies alongside high-accuracy sensors that will be utilized for the data collection and quality assurance. The sensors like Envirotech APM 411 TE, reference gas analyzers and compact modules, PMS5003, MQ-135, MICS-6814, and MOS will measure various ambient gases e.g., CO, NO₂, NH₃ particulates to monitor and allowed to improved air quality management processes.

3.2 Data pre-processing

Data Normalization

Normalization is the conversion of values measured on different scales to a conceptually equivalent scale. Min-Max normalization is a normalization technique that creates balanced value comparisons between data before and after the procedure by executing linear modifications on the original data. The following formula can be applied to this method.

$$Z_{new} = \frac{Z - \min(Z)}{\max(Z) - \min(Z)} \tag{1}$$

Max (Z) is the dataset's maximum value, Min (Z) is its minimum value, Z_{new} is the new figure obtained from the normalized results and Z = previous value

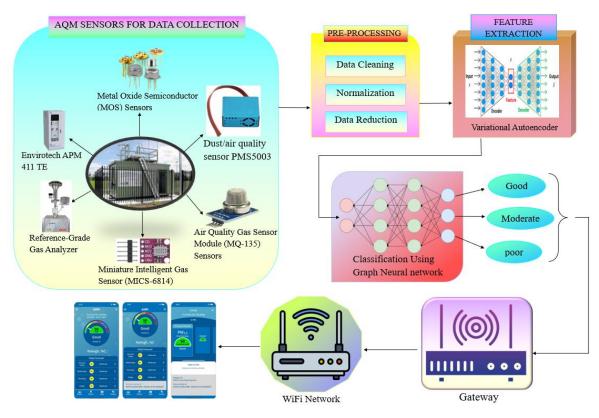


Figure 1. Proposed GNN-MBA Methodology

Data Cleaning

Data cleaning entails fixing errors, eliminating outliers, and adding missing values. Sensor datasets that contain broken sensor data and settling time zones must be cleansed. Errors in the data collection process or problems with the sensor itself are recorded as broken sensor data.

Data reduction

Data reduction technique is used to simplify large datasets by reducing their size while keeping important information. It helps improve processing speed and model performance. In this GNN-MBA it helps to remove unnecessary data, making attack detection faster and more efficient.

3.3. Feature Extraction

In feature extraction, information get from sensors is extracted from unprocessed data. The GNN-MBA technique leverages feature extraction by using VAE to capture features efficiently and raise the classification model's accuracy. VAE extracts compact and informative features from sensor outputs and encodes high-dimensional pollutant data into a lower-dimensional latent space, capture key patterns.

First, it allows to encode that extracts a latent vector from input (observation) variables and a decode that reconstructs the original variables from the latent vector.

$$x = \text{Encoder}(z) \sim q(X/z) \tag{1}$$

Here, $q(^{\chi}/_{z})$ is the decoder distribution, which models the probability of generating input data x given latent features z. Encoder(z) here refers to the decoder function in standard terminology, which takes z and tries to generate/reconstruct x.

$$\bar{z} = Decoder(x) \sim q(Z/x)$$
 (2)

Here, $q(^{\mathbb{Z}}/_{\mathcal{X}})$ is the encoder distribution, which models the probability of the latent vector z given the input x. Decoder(x) here refers to the encoder function in standard terminology which compresses input x into latent representation \overline{z} .

3.4 Classification via Graph Neural Network

This paper describes classification using GNN mechanism for air quality monitoring. GNN methods generalize neural networks to graph data. In particular, they learn feature vectors using neural networks and use them to classify data. GNNs can capture both local and global patterns, enabling them to model complex relationships in graph data effectively. GNN helps classify the air quality into three classes such as good, moderate, and poor by using its own data and the data from nearby sensors. The GNN learns to combine and update this information layer by layer.

$$g_u^{(s)} = h\left(g_u^{(s-1)}, \left\{g_v^{(s-1)} \middle| v \in N(u)\right\}\right) \tag{3}$$

Here, each of the step s, a new node presentation, $g_u^{(s)}$ is learned. Initially, $g_u^{(s-1)}$ is initialized with the node attribute

vector, and during each subsequent step, a neighbor aggregation function is applied to generate the new node representation. More specifically, common neighbor

aggregation functions for the v^{th} node. where N (u) is a set of neighboring nodes of node v.

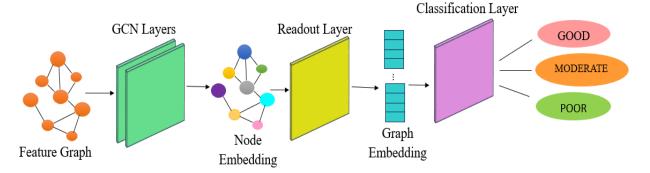


Figure 2. GNN Architecture

After completion of classification using GNN the three classes such as Good, Moderate and Poor sent to a gateway. The gateway collects and organizes the data before sending it over a Wi-Fi network to user platforms like mobile apps. Through this user can get real-time alerts, remote monitoring, scalability.

4. RESULT AND DISCUSSION

This section describes the experimental results and a performance evaluation of the suggested GNN-MBA approach using a number of assessment indicators. Python has been used in the development and implementation of the classification model. The method assesses its effectiveness using a variety of metrics including F1-score, recall, accuracy, and precision.

4.1 Data Description

The raw energy usage data was gathered from a Beijing Air Quality dataset. This dataset obtained from the UCI Machine Learning Repository, contains hourly records of meteorological data along with PM2.5 pollution measurements. Collected via the data interface released by the U.S. Embassy in Beijing and is widely used for experimental research in air quality analysis. It includes various attributes such as date, time, temperature, humidity, wind speed, wind direction, and PM2.5 concentration values, making it a comprehensive source for studying the correlation between environmental factors and air pollution levels.

4.2. Evaluation Metric

This section explains the measures that were used to evaluate the suggested GNN-MBA approach. The effectiveness of the recommended strategy has been evaluated using the F1-Score, Recall, Precision and Accuracy measures. These values are calculated as follows,

Accuracy=
$$\frac{T_e p_v + T_e N_t}{T_e p_v + F_s N_t + F_s P_v + T_e N_t}$$
(11)

Precision (PR)=
$$\frac{T_e p_v}{T_e p_v + F_s P_v}$$
 (12)

Recall (RC)=
$$\frac{T_e p_v}{T_e p_v + F_s N_t}$$
 (13)

$$F1 \text{ score} = \frac{2 \times PR \times RC}{PR + RC}$$
 (14)

4.3 Performance Analysis

According to the experimental results, the suggested GNN-MBA technique has been compared with current techniques for weather forecasting, including SARIMA, ICEEMDAN-WOA-ELM, Res-GCN-BiLSTM.

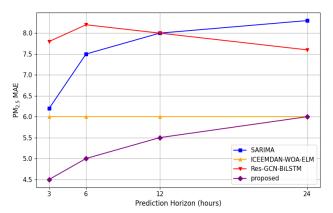


Figure 3. MAE performance

Figure.4 illustrates the MAE performance across different prediction horizons for four existing models SARIMA, ICEEMDAN-WOA-ELM, Res-GCN-BiLSTM, and the proposed method. The proposed model consistently achieves the lowest error, while SARIMA and Res-GCN-BiLSTM show higher MAE. This highlights the improved accuracy of the proposed GNN-MBA method over time.

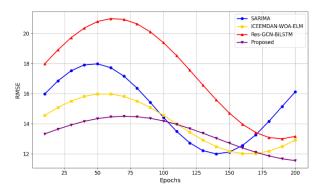


Figure 4. RMSE of proposed method

Figure.5 shows the RMSE of GNN-MBA achieves the lowest value when the epoch size is about 70 and then gradually grows when the epoch size continues to increase. The Proposed model shows consistently lower RMSE across all epochs, indicating superior performance. In contrast, Res-GCN-BiLSTM starts with the highest error and gradually improves, while ICEEMDAN-WOA-ELM and SARIMA show moderate error reduction.

The performance evaluation of classification using the Beijing Air Quality dataset is illustrated in Figure 6. For Beijing Air Quality dataset, the proposed method achieved an F1-score of 98.5%, accuracy of 97.2%, and recall of 95.4% and precision of 96.1%, demonstrating its effectiveness and reliability.

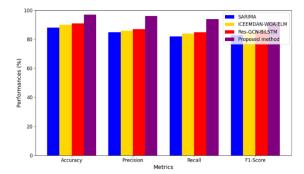


Figure 5. performance analysis

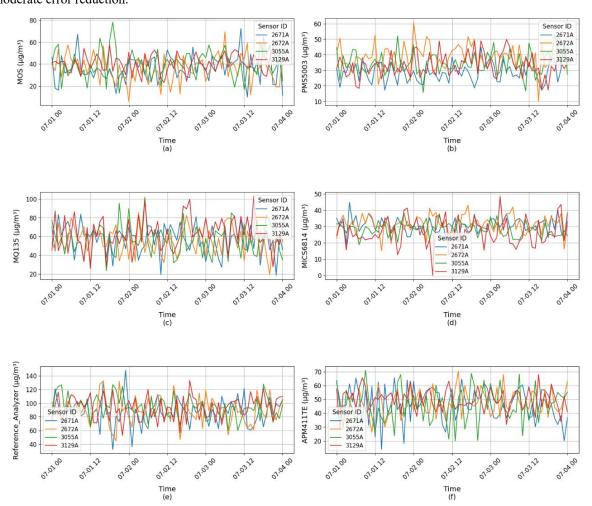


Figure 6. comparisons of observed values at monitoring station

Figure.6 shows time series plots of ozone concentration (μg/m³) recorded by six different air quality sensors—MOS, PMS5003, MQ135, MICS6814, Reference Analyzer, and APM 411 TE over three days. Each graph shows readings from four sensor units (2671A, 2672A, 3055A, 3129A),

enabling comparison of consistency, sensitivity, and performance across sensor types. The reference analyzer provides stable benchmark data, while low-cost sensors vary in responsiveness and accuracy.

5. CONCLUSION

In this paper, a novel GNN-MBA has been proposed to monitoring the air quality in real-time. The objective of the proposed GNN-MBA techniques is to provide air quality rate to save lives, improve the quality of life and protect human health through send warning messages to user mobile phones. The VAE method is utilized for feature extraction in order to efficiently capture features. The extracted features are forwards to GNN deep learning model which classifies data into three classes such as Good, Moderate and Poor, the result is then sent to the gateway and forwarded via a Wi-Fi network to the user's mobile device to get early warnings. In order to sense environmental characteristics includes temperature, humidity, pressure, and pollution levels and other harmful gases are the suggested method is contrasted with the current approaches, including the ICEEMDAN-WOA-ELM and SARIMA. The proposed GNN-MBA work achieves about 99.23% accuracy in air quality monitoring greater than existing ICEEMDAN-WOA-ELM SARIMA. Future work could be aimed at better air quality prediction by integrating satellite sensor data with lightweight models on mobile phones.

CONFLICTS OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

FUNDING STATEMENT

Not applicable.

ACKNOWLEDGEMENTS

The author would like to express his heartfelt gratitude to the supervisor for his guidance and unwavering support during this research for his guidance and support.

REFERENCES

- [1] M. Marzouk, and M. Atef, "Assessment of indoor air quality in academic buildings using IoT and deep learning", Sustainability, vol. 14, no. 12, pp. 7015, 2022. [CrossRef] [Google Scholar] [Publisher Link]
- [2] Q.A. Tran, Q.H. Dang, T. Le, H.T. Nguyen, and T.D. Le, "Air quality monitoring and forecasting system using iot and machine learning techniques", In 2022 6th International Conference on Green Technology and Sustainable Development (GTSD), pp. 786-792, 2022. [CrossRef] [Google Scholar] [Publisher Link]
- [3] C.L. Wu, H.D. He, R.F. Song, X.H. Zhu, Z.R. Peng, Q.Y. Fu, and J. Pan, "A hybrid deep learning model for regional O3 and NO2 concentrations prediction based on spatiotemporal dependencies in air quality monitoring network", *Environmental pollution*, vol. 320, pp. 121075, 2023. [CrossRef] [Google Scholar] [Publisher Link]
- [4] Y. Wei, J. Jang-Jaccard, W. Xu, F. Sabrina, S. Camtepe, and M. Boulic, LSTM-autoencoder-based anomaly detection for indoor air quality time-series data. IEEE Sensors Journal, vol. 23, no. 4, pp. 3787-3800, 2023. [CrossRef] [Google Scholar] [Publisher Link]
- [5] L. Fu, J. Li, and Y. Chen, "An innovative decision-making method for air quality monitoring based on big data-assisted artificial intelligence technique", *Journal of Innovation & Knowledge*, vol. 8, no. 2, pp. 100294, 2023. [CrossRef] [Google Scholar] [Publisher Link]

- [6] M. Ansari, and M. Alam, "An intelligent IoT-cloud-based air pollution forecasting model using univariate time-series analysis", *Arabian journal for science and engineering*, vol. 49, no. 3, pp. 3135-3162, 2024. [CrossRef] [Google Scholar] [Publisher Link]
- [7] H. Karnati, "IoT-based air quality monitoring system with machine learning for accurate and real-time data analysis", 2023. [CrossRef] [Google Scholar] [Publisher Link]
- [8] P. Asha, L.B.T.J.R.R.G.S. Natrayan, B.T. Geetha, J.R. Beulah, R. Sumathy, G. Varalakshmi, and S. Neelakandan, "IoT enabled environmental toxicology for air pollution monitoring using AI techniques", *Environmental research*, vol. 205, pp. 112574, 2022. [CrossRef] [Google Scholar] [Publisher Link]
- [9] W.A. Jabbar, T. Subramaniam, A.E. Ong, M.I. Shu'Ib, W. Wu, and M.A. De Oliveira, "LoRaWAN-based IoT system implementation for long-range outdoor air quality monitoring", *Internet of Things*, vol. 19, pp. 100540, 2022. [CrossRef] [Google Scholar] [Publisher Link]
- [10] M.N. Ramadan, M.A. Ali, S.Y. Khoo, M. Alkhedher, and M. Alherbawi, "Real-time IoT-powered AI system for monitoring and forecasting of air pollution in industrial environment", Ecotoxicology and environmental safety, vol. 283, pp. 116856, 2024. [CrossRef] [Google Scholar] [Publisher Link]
- [11] J.Á. Martín-Baos, L. Rodriguez-Benitez, R. García-Ródenas, and J. Liu, "IoT based monitoring of air quality and traffic using regression analysis", *Applied soft computing*, vol. 115, pp. 108282, 2022. [CrossRef] [Google Scholar] [Publisher Link]
- [12] Z. Liu, G. Wang, L. Zhao, and G. Yang, "Multi-points indoor air quality monitoring based on internet of things", *IEEE* access, 9, pp.70479-70492, 2021. [CrossRef] [Google Scholar] [Publisher Link]
- [13] R.R. Hemamalini, R. Vinodhini, B. Shanthini, P. Partheeban, M. Charumathy, and K. Cornelius, "Air quality monitoring and forecasting using smart drones and recurrent neural network for sustainable development in Chennai city", Sustainable Cities and Society, vol. 85, pp. 104077, 2022. [CrossRef] [Google Scholar] [Publisher Link]
- [14] D. Saravanan, and K.S. Kumar, "IoT based improved air quality index prediction using hybrid FA-ANN-ARMA model", *Materials today: proceedings*, vol. 56, pp. 1809-1819, 2022. [CrossRef] [Google Scholar] [Publisher Link]
- [15] Z. Zhang, Y. Zeng, and K. Yan, "A hybrid deep learning technology for PM2. 5 air quality forecasting", Environmental Science and Pollution Research, vol. 28, no. 29, pp. 39409-39422, 2021. [CrossRef] [Google Scholar] [Publisher Link]
- [16] D. Seng, Q. Zhang, X. Zhang, G. Chen, and X. Chen, "Spatiotemporal prediction of air quality based on LSTM neural network", *Alexandria Engineering Journal*, vol. 60, no. 2, 2021. [CrossRef] [Google Scholar] [Publisher Link]
- [17] U.A. Bhatti, M. Huang, H. Neira-Molina, S. Marjan, M. Baryalai, H. Tang, G. Wu, and S.U. Bazai, "MFFCG-Multi feature fusion for hyperspectral image classification using graph attention network", Expert Systems with Applications, vol. 229, pp. 120496, 2023. [CrossRef] [Google Scholar] [Publisher Link]
- [18] N. Jamali, M.R. Gharib, and B.O. Koma, "Neuro-Fuzzy Decision Support System for Optimization of the Indoor Air Quality in Operation Rooms", *International Journal of Robotics & Control Systems*, vol. 3, no. 1, 2023[CrossRef] [Google Scholar] [Publisher Link]
- [19] M.A. Haq, "SMOTEDNN: A novel model for air pollution forecasting and AQI classification", *Computers, Materials & Continua*, vol. 71, no. 1, 2022. [CrossRef] [Google Scholar] [Publisher Link]
- [20] Y. Wu, C. Qian, and H. Huang, "Enhanced air quality prediction using a coupled DVMD informer-CNN-LSTM model optimized with dung beetle algorithm", *Entropy*, vol.

26, no. 7, pp. 534, 2024. [CrossRef] [Google Scholar] [Publisher Link]

AUTHORS

RajuBarskar, S/o Shri J. S. Barskaris currently working as an Associate Professor in the Department of Computer Science and Engineering at of Technology, Institute Rajeev ProudhyogikiVishvidhyalaya, Bhopal (Madhya Pradesh). Dr. R. Barskar completed his Bachelor of Engineering in Information Technology from Samrat Ashok Technological Institute Vidisha, Madhya Pradesh in 2006, which is affiliated to Rajeev Gandhi ProudhyogikiVishvidhyalaya, Bhopal (Madhya Pradesh). He did Master of Technology in Computer Science and Engineering from Maulana Azad National Institute of Technology Bhopal (Madhya Pradesh) with a gold medalist in the year 2011. He also completed his Ph.D. in Computer Science and Engineering in the year 2018 from Maulana Azad National Institute of Technology Bhopal (Madhya Pradesh). He has delivered more than 05 expert lectures in different institutions which are affiliated with Rajeev Gandhi ProudhyogikiVishvidhyalaya, Bhopal (Madhya Pradesh). He has published/ presented more than 65 research papers in international/national journals and conferences. His research areas include Wireless Network, Image Processing, Machine Learning, Data Mining, VANET, String Matching Algorithms, Big Data, and High-PerformanceComputing.He is also Member of professional bodies like IEEE and ACM.



A. Vijayaraj is an Associate Professor in the Department of Information Technology, R.M.K Engineering College, RSM Nagar, Kavaraipettai, Gummidipoondi Taluk Chennai. He obtained his Bachelor's degree from Bharathidhasan University in 1997 and his Master of Engineering in CSE from Sathyabama University in 2005. He obtained his Ph.D. from Anna University Chennai. His area of specialization is Computer Networks and

communications, Mobile Computing, Information Retrieval, and Knowledge Management. He has 25 years of teaching experience from various Engineering Colleges during tenure he was Awarded Best Teacher Award thrice. He is a Member of IEEE, CSI, ISTE, IAENG, ICST, UACEE, IASTER and CSTA. He organized number of Workshops, Faculty development programs, Seminars, National and International conferences. He has written 6 books and 20 patents. He has Published 50 papers in (SCI and Scopus) reputed journals and 30 Papers published in IEEE International conferences.

Arrived: 03.03.2025 Accepted: 05.04.2025