

---

## BIBLIOGRAPHY

- A. Aarthi, P. Gayathri, N. R. Gomathi, S. Kalaiselvi, and Dr. V. Gomathi (2020), “Air Quality Prediction Through Regression Model”, *International Journal of Scientific & Technology Research*, Volume 9, Issue 03, Pp. 1-14.
- A. Samad, S. Garuda, U. Vogt, and B. Yang (2023), “Air pollution prediction using machine learning techniques – An approach to replace existing monitoring stations with virtual monitoring stations”, *Elsevier Atmospheric Environment*, Volume 310, Pp. 1-18.
- Abdelkader Dairi, Fouzi Harrou, Sofiane Khadraoui, and Ying Sun (2021), “Integrated Multiple Directed Attention-based Deep Learning for Improved Air Pollution Forecasting”, *IEEE Transactions on Instrumentation and Measurement*, Volume 70, Pp. 1-16.
- Abdellatif Bekkar, Badr Hssina, Samira Douzi, and Khadija Douzi (2021), “Air-pollution prediction in smart city, deep learning approach”, *Journal of Big Data, Springer*, Volume 8, Pp. 1-21.
- Aditya Agarwal, Vishakha Shukla, Rajesh Singh, Anita Gehlot, and Vikas Garg (2021), “Design and development of Air and Water pollution quality monitoring using IoT and Quadcopter”, *Researchgate*, Pp. 1-8.
- Ahmed Samy Moursi, Nawal El- Fishawy, Soufiene Djahel, and Marwa Ahmed Shouman (2021), “An IoT enabled system for enhanced air quality monitoring and prediction on the edge”, *Complex & Intelligent Systems, Springer*, Volume 7, Pp. 2923–2947.
- Akhil Mathew Mohan, Arathy Baby, Syam Gopi, Alan Maria George, Gopika S, and Abubeker K.M (2023), “Real-time Air Quality Index Monitoring and Alert System using IoT Technology”, *International Journal on Emerging Research Areas (IJERA)*, Volume 3, Issue 1, Pp. 1-5.
- Aman Kataria and Vikram Puri (2022), “AI- and IoT- based hybrid model for air quality prediction in a smart city with network assistance”, *Wiley, IET Networks*, Pp. 1-13.

- Amina Nazif, Nurul Izma Mohammed, Amirhossein Malakahmad, and Motasem S. Abualqumboz (2017), “Regression and multivariate models for predicting particulate matter concentration level”, *Environmental Science and Pollution Research, Springer*, Volume 25, Issue 10, Pp. 1-8.
- Amisha Gangwar, Sudhakar Singh, Richa Mishra<sup>1</sup>, and Shiv Prakash (2023), “The State-of-the-Art in Air Pollution Monitoring and Forecasting Systems using IoT, Big Data, and Machine Learning”, *Springer, Wireless Personal Communications*, Volume 130, Pp. 1699–1729.
- Anas Bushnag (2022), “An improved air quality and climate control monitoring system using fuzzy logic for enclosed areas”, *springer, Journal of Ambient Intelligence and Humanized Computing*, Pp. 1-10.
- Anikender, Kumar and Pramila Goyal (2011), “Forecasting of air quality in Delhi using principal component regression technique”, *Atmospheric Pollution Research*, Volume 2, Issue 4, Pp. 436-444.
- Anurag Barthwal and Debopam Acharya ((2021), “An IoT based Sensing System for Modeling and Forecasting Urban Air Quality”, *Springer, Wireless Personal Communications*, Volume 116, Pp. 3503–3526.
- Ayaz Hussain Bukhari, Muhammad Asif Zahoor Raja, Muhammad Shoaib, and Adiq Kausar Kiani (2022), “Fractional order Lorenz based physics informed SARFIMA-NARX model to monitor and mitigate megacities air pollution”, *Elsevier, Chaos, Solitons and Fractals*, Volume 161, Pp 1-16.
- Azim Heydari, Meysam Majidi Nezhad, Davide Astiaso Garcia, Farshid Keynia and Livio De Santoli (2022), “Air pollution forecasting application based on deep learning model and optimization algorithm”, *Clean Technologies and Environmental Policy, Springer*, Volume 24, Pp. 607–621.
- Baowei Wang, Weiwen Kong, Hui Guan, and Neal N. X (2019), “Air Quality Forecasting Based on Gated Recurrent Long Short-Term Memory Model in Internet of Things”, *IEEE Access*, Volume 7, Pp. 1-11.

- Bingchun Liu, Xiaogang Yu, Jiali Chen, and Qingshan Wang (2021), “Air pollution concentration forecasting based on wavelet transform and combined weighting forecasting model”, *Atmospheric Pollution Research, Elsevier*, Volume 12, Pp. 1-11.
- Calorine Katushabe, Santhi Kumaran, and Emmanuel Masabo (2023), “Internet of things based visualisation of effect of air pollution on the lungs using HEPA filters air cleaner”, *Heliyon*, Volume 9, Pp. 1-11.
- CanyangGuo, Genggeng Liu, and Chi-Hua Chen (2020), “Air Pollution Concentration Forecast Method Based on the Deep Ensemble Neural Network”, *Wireless Communications and Mobile Computing, Wiley*, Issue 1, Pp. 1-13.
- Chang-Yue-Shan, lin chi-yeh, Satheesh, and abimannana (2021), “Ensemble multifeatured deep learning models for air quality forecasting”, *Elsevier, Atmospheric Pollution Research*, Volume 12, Pp. 1-17.
- Claudio Badii, Stefano Bilotta, Daniele Cenni, Angelo Difino, Paolo Nesi, Irene Paoli and Michela Paolucci (2020), “High Density Real-Time Air Quality Derived Services from IoT Networks”, *Sensors*, Pp. 1-26.
- D. Hepsiba, L.M. Varalakshmi, M. Suresh Kumar, Jayasudha Subburaj, P. Vijayakumar, and S. Sheeba Rani (2021), “Automatic pollution sensing and control for vehicles using IoT technology”, *Materials Today: Proceedings, Elsevier*, Volume 45, Part 2, Pp. 3019-3021.
- D. Saravanan and K. Santhosh Kumar (2021), “Improving air pollution detection accuracy and quality monitoring based on bidirectional RNN and the Internet of Things”, *Materials Today: Proceedings, Elsevier*, Pp. 1-6.
- Dan Zhang and Simon S. Woo (2020), “Real Time Localized Air Quality Monitoring and Prediction through Mobile and Fixed IoT Sensing Network”, *IEEE Access*, Volume 8, Pp. 89584 – 89594.
- Deepak Narayan Paithankar, Abhijeet ajendra Pabale , RushikeshVilas Kolhe, P. William and Prashant Madhukar Yawalkar (2023), “Framework for implementing air quality monitoring system using LPWA-based IoT technique”, *Elsevier, Measurement: Sensors*, Volume 26, Pp. 1-6.

- Dewen Seng, Qiyang Zhang, Xuefeng Zhang, Guangsen Chen and Xiyuan Chen (2021), “Spatiotemporal prediction of air quality based on LSTM neural network”, *Knowledge-Based Systems, Elsevier*, Volume 233, Pp. 1-12.
- Dong-Her Shih, Thi Hien To, Ly Sy Phu Nguyen, Ting-Wei Wu and Wen-Ting You (2021), “Design of a Spark Big Data Framework for PM<sub>2.5</sub> Air Pollution Forecasting”, *International Journal of Environmental Research on Public Health*, Volume 18, Pp. 1-22.
- Doreswamy, Harishkumar K S, Yogesh KM and Ibrahim Gad (2020), “Forecasting Air Pollution Particulate Matter (PM<sub>2.5</sub>) Using Machine Learning Regression Models”, *Procedia Computer Science, Elsevier*, Volume 171, Pp. 2057–2066.
- Ekta Sharma, Ravinesh C. Deo, Ramendra Prasad, Alfio V. Parisi and Nawin Raj (2020), “Deep Air Quality Forecasts: Suspended Particulate Matter Modeling with Convolutional Neural and Long Short-Term Memory Networks”, *IEEE Access*, Volume 8, Pp. 209503 – 209516.
- Gabriele Casciaro, Mattia Cavaiola and Andrea Mazzino (2022), “Calibrating the CAMS European multi-model air quality forecasts for regional air pollution monitoring”, *Elsevier, Atmospheric Environment*, Volume 287, Pp. 1-15.
- Gao Huang, Chunjiang Ge, Tianyu Xiong, Shiji Song, Le Yang, Baoxian Liu, Wenjun Yin and Cheng Wu (2021), “Large scale air pollution prediction with deep convolutional networks”, *Science China Information Sciences, Elsevier*, Volume 64, Pp. 1-11.
- Ghufran Isam Drewil and Riyadh Jabbar Al-Bahadili (2022), “Air pollution prediction using LSTM deep learning and metaheuristics algorithms”, *Elsevier, Measurement: Sensors*, Volume 24, Pp.1-7.
- Gokulan Ravindiran, Gasim Hayder, Karthick Kanagarathinam, Avinash Alagumalai and Christian Sonne (2023), “Air quality prediction by machine learning models: A predictive study on the indian coastal city of Visakhapatnam”, *Elsevier, Chemosphere*, Volume 338, Pp. 1-10.

- Gordon Reikard (2019), “Volcanic emissions and air pollution: Forecasts from time series models”, *Elsevier, Atmospheric Environment*, Pp. 1-7.
- Haolin Ma, Chunjiang Zhao, Changyuan Zhai, Xiu Wang, Shuo Yang and Xiaohe Sun (2021), “Design and field experiment of precise control and monitoring system for a solid fumigant sterilizer based on IoT”, *Elsevier, Computers and Electronics in Agriculture*, Volume 189, Pp. 1-11.
- Hemanth Karnati (2023), “IoT-Based Air Quality Monitoring System with Machine Learning for Accurate and Real-time Data Analysis”, *Computer Science, Machine Learning*, Pp. 1-18.
- Himawan Nurcahyanto, Aji Teguh Prihatno, Md Morshed Alam, Md Habibur Rahman, Israt Jahan, Md Shahjalal and Yeong Min Jang (2022), “Multilevel RNN-Based PM10 Air Quality Prediction for Industrial Internet of Things Applications in Cleanroom Environment”, *Hindawi, Wireless Communications and Mobile Computing*, Volume 2022, Pp. 1-12.
- Hongmin Lia, Jianzhou Wang, Ranran Lia and Haiyan Lub (2019), “Novel analysis–forecast system based on multi-objective optimization for air quality index”, *Elsevier*, Pp. 1-39.
- Hongqian Chen, Mengxi Guan and Hui Li (2021), “Air Quality Prediction Based on Integrated Dual LSTM Model”, *IEEE Access*, Volume 9, Pp. 93285 – 93297.
- Ho-Yong Jin, Eun-Sung Jung and Duckki Lee (2020), “High-performance IoT streaming data prediction system using Spark: a case study of air pollution”, *Neural Computing and Applications*, Volume 32, Pp. 13147–13154.
- Ichrak Mokhtari, Walid Bechkit, Herve Rivano and Mouloud Riadh Yaici (2021), “Uncertainty-Aware Deep Learning Architectures for Highly Dynamic Air Quality Prediction”, *IEEE Access*, Volume 9, Pp. 14765 – 14778.
- Isura Sumeda Priyadarshana Nagahage, Ekanayaka Achchillage Ayesha Dilrukshi Nagahage and Takeshi Fujino (2021), “Assessment of the applicability of a low-cost sensor - based methane monitoring system for continuous multi- channel sampling”, *Environmental Monitoring and Assessment, Springer*, Volume 193, Pp. 1-14.

- Jagriti Saini, Maitreyee Dutta and Gonçalo Marques (2020), “Indoor Air Quality Monitoring Systems Based on Internet of Things: A Systematic Review”, *International Journal of Environmental Research on Public Health*, Volume 17, Pp. 1-21.
- Jennifer L. Moutinho, Donghai Liang, Rachel Golan, Stefanie E. Sarnat, Rodney Weber, Jeremy A. Sarnat and Armistead G. Russell (2020), “Near-road vehicle emissions air quality monitoring for exposure modeling”, *Atmospheric Environment, Elsevier*, Volume 224, Pp. 1-11.
- Jianxian Cai, Xun Dai, Li Hong, Zhitao Gao, and Zhongchao Qiu (2020), “An Air Quality Prediction Model Based on a Noise Reduction Self-Coding Deep Network”, *Mathematical Problems in Engineering, Hindawi*, Volume 2020, Pp. 1-12.
- Jianzhou Wang, Hongmin Li, Hufang Yang and Ying Wang (2021), “Intelligent multivariable air-quality forecasting system based on feature selection and modified evolving interval type-2 quantum fuzzy neural network”, *Environmental Pollution, Elsevier*, Volume 274, Pp. 1-17.
- Jing Huang, Xiaochuan Pan, Xinbiao Guo and Guoxing Li (2018), “Health impact of China’s Air Pollution Prevention and Control Action Plan: an analysis of national air quality monitoring and mortality data”, *Lancet Planet Health*, Pp. 1-11.
- Jiyang Wang, Jingrui Li and Zhiwu Li (2022), “Prediction of Air Pollution Interval Based on Data pre-processing and Multi-Objective Dragonfly Optimization Algorithm”, *Frontiers in Ecology and Evolution*, Volume 10, Pp. 1-18.
- José Ángel Martín-Baos, Luis Rodríguez-Benitez, Ricardo García-Ródenas and Jun Liu (2022) “IoT based monitoring of air quality and traffic using regression analysis” *Elsevier*, Volume 115, Pp.1-16.
- Jovan Kalajdjieski, Kire Trivodaliev, Georgina Mirceva, Slobodan Kalajdziski and Sonja Gievska (2023), “A Complete Air Pollution Monitoring and Prediction Framework”, *IEEE Access*, Volume 11, Pp. 1-15.
- Jovan Pantelic, Young Joo Son, Brant Staven and Qingyang Liu (2023), “Cooking emission control with IoT sensors and connected air quality interventions for smart and healthy homes: Evaluation of effectiveness and energy consumption”, *Elsevier, Energy and Buildings*, Volume 286, Pp. 1-14.

- Julio Buelvas, Danny M´unera, Diana P. Tob´on V, Johnny Aguirre and Natalia Gaviria (2023), “Data Quality in IoT-Based Air Quality Monitoring Systems: a Systematic Mapping Study”, *Water Air Soil Pollution*, Pp. 1-23.
- Julio H. Buelvas, Danny M´unera and Natalia Gaviria (2023), “DQ-MAN: A tool for multi-dimensional data quality analysis in IoT-based air quality monitoring systems”, *Elsevier, Internet of Things 22*, Pp.1-18.
- Jun He Yang and Chia Pang Chan (2022), “Urban Ecological Monitoring and Prediction Based on Deep Learning”, *Wireless Communications and Mobile Computing, Wiley, Volume 7*, Pp. 1-6.
- JunHo Jo, ByungWan Jo, JungHoon Kim, SungJun Kim, and WoonYong Han (2022), “Development of an IoT-Based Indoor Air Quality Monitoring Platform”, *Journal of Sensors*, Pp. 1-15.
- K. Krishna Rani Samal, Ankit Kumar Panda, Korra Sathya Babu and Santos Kumar Das (2021), “An improved pollution forecasting model with meteorological impact using multiple imputation and fine-tuning approach”, *Sustainable Cities and Society, Elsevier, Volume 70*, Pp. 1-13.
- K. Kumar and B. P. Pande (2022), “Air pollution prediction with machine learning: a case study of Indian cities”, *International Journal of Environmental Science and Technology, Springer*, Pp. 1-16.
- K. Sridhar, P. Radhakrishnan, G. Swapna, R. Kesavamoorthy, L. Pallavi and R. Thiagarajan (2023), “A modular IOT sensing platform using hybrid learning ability for air quality prediction”, *Elsevier, Measurement: Sensors, Volume 25*, Pp. 1-7.
- Kan Zheng, Shaohang Zhao, Zhe Yang, Xiong Xiong and Wei Xiang (2016), “Design and Implementation of LPWA-Based Air Quality Monitoring System”, *IEEE Access, Volume 4*, Pp. 1-8.
- Kirti Vaidya, Janhavi Hatkar, Apurva Gadekar and Shubhada Ghangare (2023), “Air Pollution Monitoring System Using IoT”, *International Research Journal of Modernization in Engineering Technology and Science, Volume:05*, Pp.1-8.

- Koel Datta Purkayastha, Ritesh Kishore Mishra, Arunava Shil and Sambhu Nath Pradhan (2021), “IoT Based Design of Air Quality Monitoring System Web Server for Android Platform”, *Wireless Personal Communications, Springer*, Volume 118, Pp. 2921–2940.
- Komal Dahekar and Prof. Rhaul Dhutire (2022), “IOT based industrial air pollution monitoring system and controlling system”, *International Research Journal of Modernization in Engineering Technology and Science*, Volume 04, Issue 06, Pp. 1-5.
- Kun Gao, Prathik Anandhan and Rakesh Kumar (2021), “Analysis and evaluation of the regional air quality index forecasting based on web-text sentiment analysis method”, *Elsevier, Environmental Impact Assessment Review*, Volume 87, Pp. 1-8.
- Liang Zhao, Huan Zhou, Rui Chen and Zhaoyang Shen (2022), “Efficient Monitoring and Adaptive Control of Indoor Air Quality Based on IoT Technology and Fuzzy Inference” *Wireless Communications and Mobile Computing*, Volume 2022, Pp. 1-14.
- Luo Zhanga, Peng Liua, Lei Zhaoc, Guizhou Wanga, Wangfeng Zhangd and Jianbo Liua (2021), “Air quality predictions with a semi-supervised bidirectional LSTM neural network”, *Elsevier, Atmospheric Pollution Research*, Volume 12, Issue 1, Pp. 328-339.
- Manal Alghieth, Raghad Alawaji, Safaa Husam Saleh and Seham Alharbi (2021), “Air Pollution Forecasting Using Deep Learning”, *International Journal of Online and Biomedical Engineering*, Volume 17, Issue 4, Pp. 1-15.
- Mannam Veera Narayana, Devendra Jalihal and S. M. Shiva Nagendra (2022), “Establishing A Sustainable Low-Cost Air Quality Monitoring Setup: A Survey of the State-of-the-Art”, *Sensors*, Volume 2022, Pp. 1-39.
- Manzoor Ansari, Mansaf Alam and Arabian (2023), “An Intelligent IoT-Cloud-Based Air Pollution Forecasting Model Using Univariate Time-Series Analysis”, *Springer, Journal for Science and Engineering*, Pp. 1-28.

- Mario Catalanoa, Fabio Galatiotob, Margaret Bellb, Anil Namdeob and Angela S. Bergantinoc (2016), “Improving the prediction of air pollution peak episodes generated by urban transport networks”, *Environmental Science & Policy, Elsevier*, Volume 60, Pp. 69-83.
- Markus Lücking, Niclas Kannengieyer, Maurice Kilgus, Till Riedel, Michael Beigl, Ali Sunyaev, and Wilhelm Stork (2020), “The Merits of a Decentralized Pollution-Monitoring System Based on Distributed Ledger Technology”, *IEEE Access*, Volume 8, Pp. 1-17.
- Martin M. Soto-Cordova, Martha Medina-De-La-Cruz and Anderson Mujaico-Mariano (2020), “An IoT based Urban Areas Air Quality Monitoring Prototype”, *International Journal of Advanced Computer Science and Applications*, Volume 11, Issue 9, Pp. 1-6.
- Mauro Castelli, Fabiana Martins Clemente, Ales Popovic, Sara Silva, and Leonardo Vanneschi (2020), “A Machine Learning Approach to Predict Air Quality in California”, *Complexity, Wiley*, Pp. 1-23.
- Middya, Asif Iqbal Roy and Sarbani (2022), “Pollutant Specific Optimal Deep Learning and Statistical Model Building for Air Quality Forecasting”, *Elsevier*, Pp. 1-14.
- Mihaita, Adriana-Simona & Dupont, Laurent & Chery, Olivier & Camargo, Mauricio & Cai, and Chen (2019), “Evaluating air quality by combining stationary, smart mobile pollution monitoring and data-driven modeling”, *Journal of Cleaner Production*, Pp. 1-55.
- Mladen Korunoski, Biljana Risteska Stojkoska and Kire Trivodaliev (2019), “Internet of Things Solution for Intelligent Air Pollution Prediction and Visualization”, *IEEE Access*, Pp. 1-7.
- Mohamed Fahim, Abderrahim El Mhouti, Tarik Boudaa and Abdeslam Jakimi (2023), “Modeling and implementation of a low-cost IoT-smart weather monitoring station and air quality assessment based on fuzzy inference model and MQTT protocol”, *Springer, Modeling Earth Systems and Environment*, Pp. 1-18.

- Mohammad A. Alolayana, Abdullah Almutairib, Suad M. Aladwania and Shiekha Alkhamees (2023), “Investigating major sources of air pollution and improving spatiotemporal forecast accuracy using supervised machine learning and a proxy”, *Elsevier, Journal of Engineering Research*, Pp. 1-7.
- Mohammed Rakib, Sanaulla Haq, Md. Ismail Hossain and Tanzilur Rahma (2022), “IoT Based Air Pollution Monitoring & Prediction System”, *Researchgate*, Pp. 1-7.
- Mohd. Tahseenul Hasan, Vijay S. Chourasia and Sanjay M. Asutkar (2019), “A Forecasting tool for Air Quality Monitoring Built up on Cloud and IoT”, *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Volume 8, Issue 10, Pp. 1-12.
- Mohieddine Benammar, Abderrazak Abdaoui, Sabbir H.M. Ahmad, Farid Touati and Abdullah Kadri (2018), “A Modular IoT Platform for Real-Time Indoor Air Quality Monitoring”, *Sensors*, Volume 18, Pp. 1-18.
- Mr. K. Azhahudurai and Dr. V. Veeramanikandan (2023), “Artificial Flora Optimization with Deep Learning Enabled Air Pollution Monitoring System in IoT Environment”, *International Journal of Intelligent Systems and Applications in Engineering IJISAE*, Volume 11(4s), Pp.22-34.
- N. Srinivasa Gupta, Yashvi Mohta, Khyati Heda, Raahil Armaan, B. Valarmathi and G. Arulkumaran (2023), “Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis”, *Hindawi, Journal of Environmental and Public Health*, Volume 2023, Pp. 1-26.
- Nabil Mohamed Eldakhly, Magdy Aboul-Ela and Areeg Abdalla (2018), “A Novel Approach of Weighted Support Vector Machine with Applied Chance Theory for Forecasting Air Pollution Phenomenon in Egypt”, *International Journal of Computational Intelligence and Applications*, Volume 17, Issue 2, Pp. 1-29.
- Naser Hossein Motlagh, Pranvera Kortoci, Xiang Su, Lauri Lovén, Hans Kristian Hoel, Sindre Bjerkestrand Haugsvær, Varun Srivastava, Casper Fabian Gulbrandsen, Petteri Nurmi, and Sasu Tarkoma (2023), “Unmanned Aerial Vehicles for Air Pollution Monitoring: A survey”, *IEEE Internet of Things Journal*, Pp. 1-20.

- Ning Jin, Yongkang Zeng, Ke Yan and Zhiwei Ji (2021), “Multivariate Air Quality Forecasting with Nested Long Short-Term Memory Neural Network”, *IEEE Transactions on Industrial Informatics*, Volume 17, Issue 12, Pp. 8514 – 8522.
- Olutosin Taiwo and Absalom E. Ezugwu (2021), “Internet of Things-Based Intelligent Smart Home Control System”, *Hindawi, Security and Communication Networks*, Volume 7, Pp.1-17.
- Oscar Alvear, Nicola Roberto Zema, Enrico Natalizio and Carlos T. Calafate (2017), “Using UAV-Based Systems to Monitor Air Pollution in Areas with Poor Accessibility”, *Hindawi, Journal of Advanced Transportation*, Volume 2017, Pp. 1-14.
- P. Chourey, N. J. Singh, K. Soni and R. Agarwal (2022), “SODAR Based Meteorological Sensor Network for Air Pollution Monitoring in Northern India”, *MA\_PAN-Journal of Metrology Society of India, Springer*, Pp. 1-22.
- Pau Ferrer-Cid, Jose M. Barcelo-Ordinas and Jorge Garcia-Vidal (2021), “Graph Learning Techniques Using Structured Data for IoT Air Pollution Monitoring Platforms”, *IEEE Access*, Pp. 1-12.
- Philipp Hähnel, Jakub Marešček, Julien Monteil and Fearghal O’ Donncha (2020), “Using deep learning to extend the range of air pollution monitoring and forecasting”, *Journal of Computational Physics, Elsevier*, Volume 408, Pp. 1-13.
- Prashant Kumar, Hamid Omidvarborna and Runming Yao (2023), “A parent-school initiative to assess and predict air quality around a heavily trafficked school”, *Elsevier, Science of the Total Environment*, Volume 86, Pp. 1-13.
- Pritthijit Nath, Pratik Saha, Asif Iqbal Middy and Sarbani Roy (2021), “Long-term time-series pollution forecast using statistical and deep learning methods”, *Neural Computing and Applications, Springer*, Volume 33, Pp. 12551–12570.
- Puneet Kalia and Mamta Alam Ansari (2020), “IOT based air quality and particulate matter concentration monitoring system”, *Elsevier, Materials Today: Proceedings*, Pp. 1-8.

- Q. P. Ha, S. Metia, and M. D. Phung (2020), “Sensing Data Fusion for Enhanced Indoor Air Quality Monitoring”, *IEEE Access*, Pp.1-12.
- Qi Zhang, Yang Han, Victor O. K. Li and Jacqueline C. K. Lam (2022), “Deep-AIR: A Hybrid CNN-LSTM Framework for Fine-Grained Air Pollution Estimation and Forecast in Metropolitan Cities”, *IEEE Access*, Volume 10, Pp. 55818 – 55841.
- Qing Tao, Fang Liu and Yong Denis Sidorov (2019), “Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU”, *IEEE Access*, Volume 7, Pp. 76690 – 76698.
- Qing Zhou, Minghua Xiao, Lei Lu, Jun Zeng, Wenting He, Chao Li and Yulun Shi (2022), “A Data-Secured Intelligent IoT System for Agricultural Environment Monitoring”, *Hindawi, Wireless Communications and Mobile Computing*, Volume 2022, Pp. 1-12.
- Qingchun Guo, Zhenfang He, Shanshan Li, Xinzhou Li, Jingjing Meng, Zhanfang Hou, Jiazhen Liu and Yongjin Chen (2020), “Air Pollution Forecasting Using Artificial and Wavelet Neural Networks with Meteorological Conditions”, *Aerosol and Air Quality Research*, Volume 20, Issue 6, Pp. 1-11.
- R. Senthilkumar, P. Venkatakrisnan, and N. Balaji (2020), “Intelligent Based Novel Embedded System Based IoT Enabled Air Pollution Monitoring System”, *Elsevier*, Volume 77, Pp. 1-29.
- R. Smit, P. Kingstona, D.W. Nealea, M.K. Brown, B. Verrana and T. Nolana (2019), “Monitoring on-road air quality and measuring vehicle emissions with remote sensing in an urban area”, *Atmospheric Environment, Elsevier*, Volume 218, Pp. 1-13.
- Rajnish Rakholia, Quan Le, Bang Quoc Ho, Khue Vu and Ricardo Simon Carbajom (2023), “Multi-output machine learning model for regional air pollution forecasting in Ho Chi Minh City, Vietnam”, *Elsevier, Environment International*, Volume 173, Pp. 1-13.
- Ranran Li, Yuqi Dong, Zhijie Zhu, Chen Li and Hufang Yang (2019), “A dynamic evaluation framework for ambient air pollution monitoring”, *Elsevier, Applied Mathematical Modelling*, Volume 65, Pp.52–71.

- Raquel Espinosa, José Palma, Fernando Jiménez, Joanna Kamińska, Guido Sciavicco and Estrella Lucena-Sánchez (2021), “A time series forecasting based multi-criteria methodology for air quality prediction”, *Applied Soft Computing, Elsevier*, Volume 13, Pp. 1-25.
- Reginald Ekene Ogu, Gloria A. Chukwudebe, Ifeyinwa E. Achumba and Nkwachukwu Chukwuchekwa (2022), “A Robust IoT-based Air Quality Monitoring Node for Multi-Location Deployment”, *International Journal of Engineering Research & Technology (IJERT)*, Volume 11, Issue 03, Pp. 1-6.
- Rodrigo de Medrano, Víctor de Buen Remiro and Jos’e L. Aznarte (2021), “SOCAIRE: Forecasting and monitoring urban air quality in Madrid”, *Elsevier, Environmental Modelling and Software*, Volume 143, Pp.1-20.
- Rui Dong, Raymond Fisman, Yongxiang Wang, and Nianhang Xu (2021), “Air Pollution, Affect, and Forecasting Bias: Evidence from Chinese Financial Analysts”, *Elsevier, Journal of Financial Economics*, Volume 139, Issue 3, Pp. 971-984.
- S Abirami and P Chitra (2021), “Regional air quality forecasting using spatiotemporal deep learning”, *Journal of Cleaner Production, Elsevier*, Volume 283, Pp. 1-14.
- S. Metia , Q.P. Ha, H.N. Duc and Y. Scorgie (2020), “Urban Air Pollution Estimation Using Unscented Kalman Filtered Inverse Modeling With Scaled Monitoring Data”, *Elsevier, Sustainable Cities and Society*, Volume 54, Pp.1-27.
- S.Chenchu Jyoshna, S.Chandrakala and P.Srividya (2021), “IoT Based Air Quality Monitoring System”, *Dogo Rangsang Research Journal*, Volume 08, Issue 14, Pp. 1-5.
- S.M.S.D. Malleswari and T. Krishna Mohana (2022), “Air pollution monitoring system using IoT devices: Review”, *Elsevier, Materials Today: Proceedings*, Volume 51, Part 1, Pp.1147-1150.
- Saba Gul, Gul Muhammad Khan and Sohail Yousaf (2022), “Multi- step short- term PM2.5 forecasting for enactment of proactive environmental regulation strategies”, *Environmental Monitoring and Assessment, Springer*, Volume 194, Issue 386, Pp. 1-14.

- Said Munira, Martin Mayfielda, Daniel Cocab and Stephen A. Jubba (2019), “Structuring an integrated air quality monitoring network in large urban areas – Discussing the purpose, criteria and deployment strategy”, *Elsevier, Atmospheric Environment: X* 2, Volume 2019, Pp. 1-13.
- Saman Taheri and Ali Razban (2021) , “Learning-based CO2 concentration prediction: Application to indoor air quality control using demand-controlled ventilation”, *Building and Environment* , Volume 205, Pp. 1-12.
- Sami Kaivonen and Edith C.-H. Ngai (2020), “Real-time air pollution monitoring with sensors on city bus”, *Digital Communications and Networks*, Volume 6, Pp. 23–30.
- Santanu Metia, Huynh A. D. Nguyen and Quang Phuc Ha (2021), “IoT-Enabled Wireless Sensor Networks for Air Pollution Monitoring with Extended Fractional-Order Kalman Filtering”, *Sensors*, Volume 21, Pp. 1-16.
- Saverio De Vito, Girolamo Di Francia, Elena Esposito, Sergio Ferlito, Fabrizio Formisano and Ettore Masseraa (2020), “Adaptive machine learning strategies for network calibration of IoT smart air quality monitoring devices” *Elsevier*, Volume 136, Pp. 264-271.
- Shajulin Benedict, Rumaize P. and Jaspreet Kaur (2020), “IoT Blockchain Solution for Air Quality Monitoring In Smartcities”, *IEEE Access*, Pp. 1-9.
- Shan Jiang, Zu-Guo Yu, Vo V. Anh and Yu Zhou (2021), “Long- and short-term time series forecasting of air quality by a multiscale framework”, *Environmental Pollution, Elsevier*, Volume 271, Pp. 1-10.
- Shengdong Du, Tianrui Li, Yan Yang and Shi-Jinn Horng (2021), “Deep Air Quality orecasting Using Hybrid Deep Learning Framework”, *IEEE Transactions on Knowledge and Data Engineering*, Volume 33, Issue 6, Pp. 2412 – 2424.
- Silvia Liberata Ullo and Ganesh Ram Sinha, “Advances in Smart Environment Monitoring Systems Using IoT and Sensors”, *Sensors 2020*, Pp. 1-18.

- Subin Han, Yongmi Park, Namkyu Noh, Joo-Hong Kim, Jae-Jin Kim, Baek-Min Kim and Wonsik Choi (2023), “Spatiotemporal variability of the PM<sub>2.5</sub> distribution and weather anomalies during severe pollution events: Observations from 462 air quality monitoring stations across South Korea”, *Elsevier, Atmospheric Pollution Research*, Volume 14, Pp.1-12.
- T Dineshkumar, V Suresh Babu, Pachaivannan Partheeban, and R Puviarasi (2021), “Air Quality Monitoring System Based on IoT”, *Journal of Physics: Conference Series 1*, Pp. 1-7.
- T. Mandava, S. Chen, O. Isafiade and A. Bagula (2020), “An IoT Middleware for Air Pollution Monitoring in Smart Cities: A Situation Recognition Model”, *Researchgate*, Pp. 1-20.
- TaofeekDolapo Akinosho, Muhammad Bilal, EndaThomas Hayes, Anuoluwapo Ajayi, Ashraf Ahmed and Zaheer Khan (2023), “Deep learning-based multi-target regression for traffic-related air pollution forecasting”, *Elsevier, Machine Learning with Applications*, Volume 12, Pp.1-28.
- Tarun Kumara and Amulya Doss (2023), “AIRO: Development of an Intelligent IoT-based Air Quality Monitoring Solution for Urban Areas”, *Elsevier, Procedia Computer Science*, Volume 218, Pp.262–273.
- Tong Liu, Alexis K. H. Lau, Kai Sandbrink, and Jimmy C. H. Fung (2018), “Time Series Forecasting of Air Quality Based on Regional Numerical Modeling in Hong Kong”, *Journal of Geophysical Research: Atmospheres*, Volume 123, Issue 3, Pp. 1-23.
- Uddesh U Naik, Shreyash R Salgaokar and Shreeyank Jambhale (2023), “IoT Based Air Pollution Monitoring System”, *International Journal of Scientific Research & Engineering Trends*, Volume 9, Issue 3, Pp.1-4.
- Uzair Aslam Bhatti, Yuhuan Yan, Mingquan Zhour, Sajid Ali, Aamir Hussain, Huo Qingsong, Zhaoyuan Yu and Lingwang Yuan (2021), “Time Series Analysis and Forecasting of Air Pollution Particulate Matter (PM<sub>2.5</sub>): An SARIMA and Factor Analysis Approach”, *IEEE Access*, Volume 9, Pp. 41019 - 41031.

- V. N. Kukre, Divyashree More, Ojus Thopate, Aarya Nigade and Sahil Pilke (2023), "Air Quality Monitor System Using IoT And Machine Learning", *International Research Journal of Modernization in Engineering Technology and Science*, Volume 05, Issue 03, Pp.1-4.
- Waheb A. Jabbar, Thanasrii Subramaniam, Andre Emelio Ong, Mohd Iqmal Shu'lb, Wenyan Wu, Mario A. de Oliveira (2022), "LoRaWAN-Based IoT System Implementation for Long-Range Outdoor Air Quality Monitoring", *Elsevier, Internet of Things*, Volume 19, Pp.1-25.
- WeilinWang, Wenjing Mao, Xueli Tong and Gang Xu (2021), "A Novel Recursive Model Based on a Convolutional Long Short-Term Memory Neural Network for Air Pollution Prediction", *Remote Sensor*, Volume 13, Pp. 1-20.
- Wen Tsai Sung and Sung Jung Hsiao (2021), "Building an indoor air quality monitoring system based on the architecture of the Internet of Things", *Journal on Wireless Communications and Networking, Springer*, Pp. 1-40.
- Wenjing Mao, Limin Jiao, Weilin Wang, Jianlong Wang, Xueli Tong and Suli Zhao (2021), "A hybrid integrated deep learning model for predicting various air pollutants", *GIScience & Remote Sensing*, Volume 58, Issue 8, Pp.1395-1412.
- Xiangyu Zou, Jinjin Zhao, Duan Zhao, Bin Sun, Yongxin He and Stelios Fuentes (2021), "Air Quality Prediction Based on a Spatiotemporal Attention Mechanism", *Mobile Information Systems, Hindawi*, Volume 2021, Pp. 1-12.
- Xu Fenga, Tzung-May Fua, Hansen Caoa, Heng Tiana, Qi Fanb and Xiaoyang Chenb (2019), "Neural network predictions of pollutant emissions from open burning of crop residues: Application to air quality forecasts in southern China", *Elsevier, Atmospheric Environment*, Volume 204, Pp. 22–31.
- Yanghao Ye, Qing Wang and Jian Wang (2021), "Green city air monitoring and architectural digital art design based on IoT embedded system", *Elsevier, Environmental Technology & Innovation*, Volume 23, Pp. 1-14.
- Yanlai Zhou, Fi-John Chang, Li-Chiu Chang, I-Feng Kao and Yi-Shin Wang (2019), "Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts", *Elsevier, Journal of Cleaner Production*, Volume 209, Pp. 134-145.

- Yanzhao Li Ju-e Guo, Shaolong Sun, Jianing Li, Shouyang Wang and Chengyuan Zhang (2022), “Air quality forecasting with artificial intelligence techniques: A scientometric and content analysis”, *Elsevier, Environmental Modelling and Software*, Volume 149, Pp.1-17.
- Ying Zhang, Rongrong Zhang, Qunfei Ma, Yanhao Wang, Qingqing Wang, Zihao Huang and Linyan Huang (2020), “A feature selection and multi-model fusion-based approach of predicting air quality”, *ISA Transactions, Elsevier*, Volume 100, Pp. 210-220.
- Ying Zhang, Yanhao Wang, Minghe Gao, Qunfei Ma, Jing Zhao, Rongrong Zhang, Qingqing Wang, and Linyan Huang (2019), “A Predictive Data Feature Exploration-Based Air Quality Prediction Approach”, *IEEE Access*, Volume 7, Pp.1-12.
- Yongliang Feng (2022), “Air Quality Prediction Model Using Deep Learning in Internet of Things Environmental Monitoring System”, *Hindawi, Mobile Information Systems*, Volume 2022, Pp. 1-9.
- Yu Huang, Josh Jia-Ching Ying and Vincent S. Tseng (2021), “Spatio-attention embedded recurrent neural network for air quality prediction”, *Knowledge-Based Systems, Elsevier*, Volume 233, Pp.1-14.
- Yuan Huang, Yuxing Xiang, Ruixiao Zhao and Zhe Cheng (2020), “Air Quality Prediction Using Improved PSO-BP Neural Network”, *IEEE Access*, Volume 8, Pp. 99346 – 99353.
- Yue-Shan Chang, Hsin-Ta Chiao, Satheesh Abimannan, Yo-Ping Huang, Yi-Ting Tsaia and Kuan-Ming Lina (2020), “An LSTM-based aggregated model for air pollution forecasting”, *Atmospheric Pollution Research, Elsevier*, Volume 11, Issue 8, Pp. 1451-1460.
- Yu-Lin Zhao, Jiali Tang, Han-Pang Huang, Ze Wang, Tse-Lun Chen, Chih-Wei Chiang and Pen-Chi Chiang (2020), “Development of IoT Technologies for Air Pollution Prevention and Improvement”, *Aerosol and Air Quality Research*, Volume 20, Pp. 2874–2888.

- Yuting Yang, Gang Mei and Stefano Izzo (2022), “Revealing Influence of Meteorological Conditions on Air Quality Prediction Using Explainable Deep Learning”, *IEEE Access*, Volume 10, Pp. 50755 – 50773.
- Zeba Idrees and Lirong Zheng (2019), “Low Cost Air Pollution Monitoring Systems: A Review of the Protocols and the Enabling Technologies”, *Journal of Industrial Information Integration*, Pp.1-12.
- Zhibin Liu, Guangwen Wang, Liang Zhao, and Guangfei Yang (2021), “Multi-Points Indoor Air Quality Monitoring Based on Internet of Things” *IEEE*, Volume 9, Pp. 70479 – 70492.
- Zhihe Zhao, Jiaheng Wang, Chenxu Fu, Zhenbang Liu, Dawei Liu and Bailiang Li (2018), “Design of a Smart Sensor Network System for Real-Time Air Quality Monitoring on Green Roof”, *Hindawi, Journal of Sensors*, Volume 2018, Pp. 1-13.
- Zixi Zhao, Jinran Wu, Fengjing Cai, Shaotong Zhang and You-Gan Wang (2022), “A statistical learning framework for spatial-temporal feature selection and application to air quality index forecasting”, *Elsevier, Ecological Indicators*, Volume 144, Pp. 1-16.

---

## PUBLICATIONS

### International Journals:

1. M. Dhanalakshmi & Dr. V. Radha. (2023), “Novel Regression and Least Square Support Vector Machine Learning Technique for Air Pollution Forecasting”, SSRG International Journal of Engineering Trends and Technology, Vol. 71, Issue. 04, pp: 147-158, ISSN: 2231-5381. **(Scopus Indexed)**
2. M. Dhanalakshmi & Dr. V. Radha. (2022), “Discretized Linear Regression and Multiclass Support Vector based Air pollution Forecasting Technique”, SSRG International Journal of Engineering Trends and Technology, Vol. 70, Issue. 11, pp: 315-323, ISSN: 2231-5381. **(Scopus Indexed)**
3. M. Dhanalakshmi & Dr. V. Radha. (2021), “Vehicular Air Purifier – IoT Enabled System with Artificial Intelligence to Prevent Air Pollution”, International Research Journal on Advance Science Hub, Vol. 03, Issue.07S, pp: 75-78, ISSN: 2582-4376. **(peer-reviewed)**
4. M. Dhanalakshmi & Dr. V. Radha (2022), “Air Pollution Forecasting using Novel Feature Selection and Classification Model”, International Journal of Engineering Technology and Management Sciences, Vol. 06, Issue. 6, pp.: 20-27, ISSN: 2581-4621. **(peer-reviewed)**
5. M. Dhanalakshmi & Dr. V. Radha. (2021), “A Survey Paper on Vehicles Emitting Air Quality and Prevention of Air Pollution by using IoT along with Machine Learning Approaches”, Turkish Journal of Computer and Mathematics Education, Vol. 12, No. 11, pp: 5950-5962, ISSN: 1309-4653. **(peer-reviewed)**

### Book Chapters and International Conference:

1. M. Dhanalakshmi & Dr. V. Radha, (2022),” Enhanced pre-processing Technique for Air Pollution Forecasting System Using Big Data and Internet of Things “, Data Intelligence and Cognitive Informatics, ISBN 978-981-19-6003-1, DOI: 10.1007/978-981-19-6004-8. **(Springer – Scopus Indexed)**

2. M. Dhanalakshmi & Dr. V. Radha, (2022), “Comparative Study on IoT based System for Enhancing Air Pollution Forecasting Accuracy and Air Quality Index Monitoring”, International Conference on Emerging Trends in IoT and Computing Technologies. ISBN 9781003350057. DOI: 10.1201/9781003350057 (**CRC Press Taylor and Francis Group –Scopus Indexed**)
  
3. M. Dhanalakshmi & Dr. V. Radha, (2022), “Novel Regression and Feature Selection Techniques for Air Pollution Forecast with IoT”, Second International Conference on Emerging Trends on commerce, Science, Engineering, Management and Technology (ICCSEMT - II), pp:109.(**Conference Proceedings**)



**Avinashilingam Institute for Home Science and Higher Education for Women**

(Deemed to be University Estd. u/s 3 of UGC Act 1956, Category 'A' by MHRD  
Re-accredited with A++ Grade by NAAC. CGPA 3.65/4, Category I by UGC  
Coimbatore - 641 043, Tamil Nadu, India

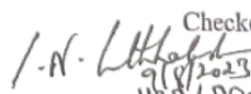
**Appendix L2**

**(Item No 5 of  
Check List) Details of Research  
Publications**

S.No	Article	Journal	Other Details Vol/No/Page No/ Year	Published in UGC- CARE / Scopus Indexed/ Web of Science
1	Discretized Linear Regression and multiclass Support Vector Based Air Pollution Forecasting Technique	SSRG International Journal of Engineering Trends and Technology	Vol : 70 Issue 11, 315-323 November 2022	Scopus ISSN: 2231-5381
2	Novel Regression and Least Square Support Vector Machine Learning Technique for Air Pollution Forecasting	SSRG International Journal of Engineering Trends and Technology	Vol: 71 Issue 4, 147-158 April 2023	Scopus ISSN: 2231-5381

\*Proof of list of Journals from Internet to be attached along with copies of reprints.

Scholar :   
Supervisor :   
08/08/23

Checked By:  
  
9/8/2023  
HOD/Dean of Respective School

The scholar Mrs. Dhanalakshmi, M (19PHCESFECT)  
has published her papers in the following journal:

1. SSRG International Journal of Engineering Trends  
and Technology - is indexed and active in Scopus from  
2019 to present. The scholar published her paper in  
~~Vol. 70, No. 11, November, 2022 Pg. 315 - 323 and~~
2. Vol. 71, No. 4, April, Pg. 147 - 158, 2023.

J. J. Gill  
08.08.23.

Original Article

# Discretized Linear Regression and Multiclass Support Vector Based Air Pollution Forecasting Technique

M. Dhanalakshmi<sup>1</sup>, V. Radha<sup>2</sup>

<sup>1,2</sup>Department of Computer Science, Avinashilingam Institute of Home Science and Higher Education for Women, Coimbatore.

<sup>1</sup>Corresponding Author : [dhanalaxmi2289@gmail.com](mailto:dhanalaxmi2289@gmail.com)

Received: 16 August 2022

Revised: 11 November 2022

Accepted: 17 November 2022

Published: 26 November 2022

**Abstract** - Air pollution is a vital issue emerging from the uncontrolled utilization of traditional energy sources as far as developing countries are concerned. Hence, ingenious air pollution forecasting methods are indispensable to minimize the risk. To that end, this paper proposes an Internet of Things (IoT) enabled system for monitoring and controlling air pollution in the cloud computing environment. A method called Linear Regression and Multiclass Support Vector (LR-MSV) IoT-based Air Pollution Forecast is proposed to monitor the air quality data and the air quality index measurement to pave the way for controlling effectively. Extensive experiments carried out on the air quality data in the India dataset have revealed the outstanding performance of the proposed LR-MSV method when benchmarked with well-established state-of-the-art methods. The results obtained by the LR-MSV method witness a significant increase in air pollution forecasting accuracy by reducing the air pollution forecasting time and error rate compared with the results produced by the other state-of-the-art methods.

**Keywords** - Internet of Things, Cloud Computing, Wavelet, Sliding Window, Linear Regression, Correlation, Multiclass, Support Vector.

## 1. Introduction

World health organization (WHO) disclosed that air pollution is highly susceptible to the sky-scraping environmental hazard to health and has resulted in a high mortality rate. Future research developments require machine learning techniques to predict air pollution quality monitoring and control. In addition to air pollution detection, the accuracy of quality being monitored has to be concentrated. This objective can be arrived at by examining the environment via IoT and reshaping neural networks.

## 2. Methodology

Nowadays, air pollution monitoring and control are used for numerous tasks measuring wind speed and direction, monitoring vehicle emissions etc. Therefore air pollution monitoring and control have to be performed in the earlier stage to control the hazard caused to humans, forests and animals. This air pollution monitoring and control issue is obviously addressed by crafting mechanisms by capturing the air quality data from various sensors (i.e., internet of things), storing and processing via a Cloud Computing environment. Air pollution monitoring and control via attention-based scheme and bidirectional RNN have been proposed in the recent literature

Despite satisfactory results provided by both classes of methods, both time and frequency aspects subdue noisy features and, owing to complicated data, compromise both accuracy and error rate. To address the above-said issues, in

this work, a Linear Regression and Multiclass Support Vector (LR-MSV) IoT-based Air Pollution Forecast method for air pollution monitoring and control is proposed. Figure 1 shows the block diagram of the LR-MSV methodology.

As shown the above figure1, the LR-MSV method includes three stages. First, pre-processing of the air quality data in India to obtain noise-reduced pre-processed data is made computationally efficient by employing Wavelet Sliding Window-based Multi-resolute Pre-processing model.

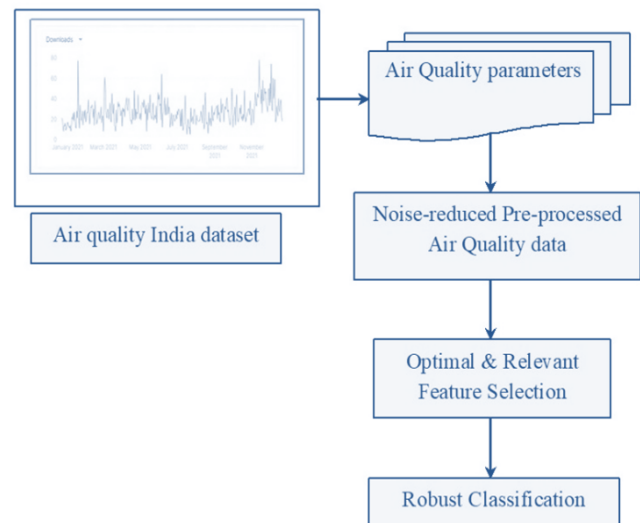


Fig. 1 Block diagram of LR-MSV method



Second, with the noise-reduced pre-processed air quality data, optimal and relevant features for further processing are made using Linear Regression and a Correlation-based Feature Selection model

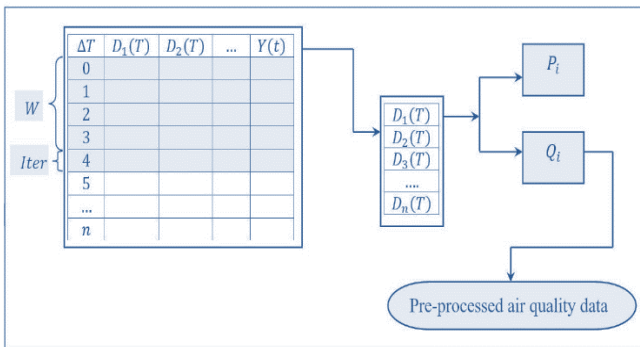
Finally, with the identified relevant, a machine learning technique is applied to monitor air quality data through Air Quality Index efficiently and accordingly place the results in the AQI\_bucket via multiclass support vectors. A detailed description of the proposed LR-MSV method is presented in the forthcoming sections, following a description of the Air Pollution Monitoring and Control System model.

**2.1. Wavelet Sliding Window-based Multi-resolute Pre-Processing Model**

Air pollution (AP) influences not only the environment but the body parts of humans and the respiratory system. Hence, systematic Air Quality (AQ) monitoring and control is required to evaluate the AP level and anticipate the pollutant concentrations with minimal noise. The wavelet decomposition model is an extensively utilized signal processing method in time series prediction. Its fundamental postulate remains in decomposing a non-smooth discrete time series air quality data into a blend of progressions with numerous high-frequency feature components ‘H’ and low-frequency coarse-grained component ‘L’.

Here, the frequency of high-frequency feature components depends on the number of layers of the wavelet decomposition.

In this work, the Wavelet Sliding Window-based Multi-resolute Pre-processing model is applied to the raw air quality data obtained from the respective sensors to address the aspects involving both time and frequency. The multi-resolute here refers to the time and frequency while performing pre-processing via sliding window towards noise removal. The sliding window establishes per sample for each time instance ‘T’ and the samples of ‘T<sub>n</sub>’ utilizes the values as ‘[T<sub>n-WS</sub>, T<sub>n</sub>]’, where ‘T<sub>n</sub>’ records the air quality values and window size ‘WS’ respectively. Figure 2 shows the structure of the Wavelet Sliding Window-based Multi-resolute Pre-processing model.



**Fig. 2 Wavelet Sliding Window-based Multi-resolute Pre-processing model**

As shown from the above figurative representation, the sliding window size value influences the number of time series samples and features present in the dataset. Owing to this, a small sliding window size refers to more air quality samples, whereas a larger sliding window size refers to fewer samples and more features. This sliding window model is then initially formulated as given below.

$$y(W + Iter) = f(T_0, \dots, T_W, D_{i,0}, \dots, D_{i,W}, T_{W+Iter}, Y_{W+Iter}) \tag{1}$$

$$f(T_0, \dots, T_W, D_{i,0}, \dots, D_{i,W}, Y_0, \dots, Y_W, T_{W+Iter}, Y_{W+Iter}) \tag{2}$$

In the sliding window, wavelet decomposition is utilized by the cloud server upon recording each air quality data from the corresponding sensors (i.e., raw air quality time series data) that discretize high-frequency signal with high-frequency feature components from the low-frequency coarse-grained components to acquire more data features. The decomposition process for each feature is mathematically formulated as given below.

$$P_{i+1} = L(P_i) \tag{3}$$

$$Q_i = H(Q_i) \tag{4}$$

From the above two equations (3) and (4), ‘P<sub>i</sub>’, ‘Q<sub>i</sub>’ denotes the low-frequency coarse-grained component and high-frequency feature component, respectively, with ‘L’ forming the low pass filter and ‘H’ representing the high pass filter. During wavelet transform based pre-processing, each layer of the decomposed signal present in the sliding window is the bisection of the pre-decomposed air quality signal data. Hence, dual interpolation reconstructions are required to retrieve the signal length (i.e., pre-processed air quality data) and the reconstruction formula performed by the cloud server for each sensor is mathematically formulated as given below.

$$P_i = (L_2)^i P_i \tag{5}$$

$$Q_i = (L_2)^{i-1} H_2 Q_i \tag{6}$$

From the above two equations (5) and (6), ‘L<sub>2</sub>’ and ‘H<sub>2</sub>’ denotes the dual operators with which the cloud server distinguishes clearly between noise and significant information for air quality data at an hourly and daily level of various stations across multiple cities in India. The pseudocode representation is given below.

As given in the above Time and Frequency-based Sliding Window Pre-processing algorithm, the objective remains in returning the pre-processed air quality data by the cloud server for each corresponding sensor with minimum error; with this objective, initially sliding window for the respective sensors (i.e., device extracting air quality data like date, PM2.5, PM10, NO and so on) are modelled. Second, wavelet decomposition is formulated for the respective sensors according to low-frequency and high-frequency feature components. Third,

with the dual operators, the cloud server returns the noise-reduced pre-processed air quality data for further processing.

**Algorithm. 1. Time and Frequency-based Sliding Window Pre-processing**

<p><b>Input:</b> Dataset ‘<i>DS</i>’, Cloud Server ‘<i>CS</i>’, IoT Devices or Sensors ‘<math>S = S_1, S_2, \dots, S_n</math>’, Features ‘<math>F = F_1, F_2, \dots, F_n</math>’, Air Quality data ‘<math>D = D_1, D_2, \dots, D_n</math>’</p>
<p><b>Output:</b> Noise reduced pre-processed air quality data ‘<i>PD</i>’</p>
<p>Step 1: <b>Initialize</b> time instance ‘<i>T</i>’                  Step 2: <b>Begin</b>                  Step 3: <b>For</b> each Dataset ‘<i>DS</i>’ (Air Quality data ‘<math>D = D_1, D_2, \dots, D_n</math>’) with Cloud Server ‘<i>CS</i>’ and IoT Devices or Sensors ‘<math>S = S_1, S_2, \dots, S_n</math>’                  Step 4: Formulate a sliding window as in equations (1) and (2)                  Step 5: <b>For</b> each Features ‘<i>F</i>’                  Step 6: Perform decomposition as in equations (3) and (4)                  Step 7: Model dual interpolation reconstructions to retrieve pre-processed air quality data as in equations (5) and (6)                  Step 8: <b>Return</b> pre-processed air quality data ‘<i>PD</i>’                  Step 9: <b>End for</b>                  Step 10: <b>End for</b>                  Step 11: <b>End</b></p>

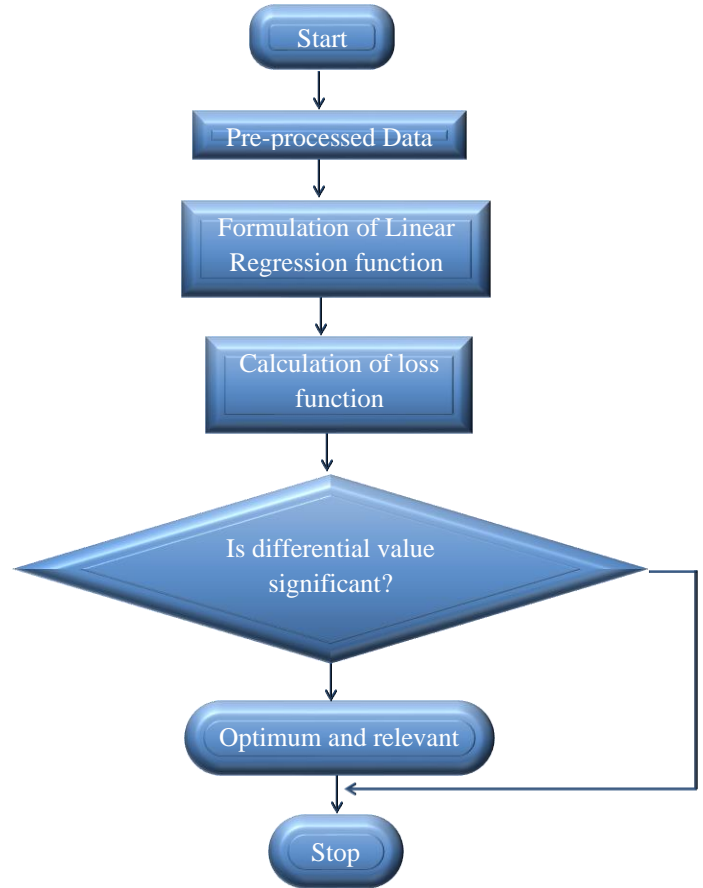
**2.2. Linear Regression and Correlation-based Feature Selection**

With the pre-processed air quality data, the next step remains to select the optimal features. Prevailing feature selection models like Monte Carlo [2] take no notice of the dependencies between the pre-processed features. A new model named Linear Regression and Correlation-based Feature Selection has been proposed based on the correlation between the features. Figure 4 shows the flow diagram of the Linear Regression and Correlation-based Feature Selection model.

As shown in the figure, let us assume that the independent pre-processed air quality data is ‘ $PD = (PD_1, PD_2, \dots, PD_n)$ ’ with regression coefficients are ‘ $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ ’, then, the Linear Regression and Correlation-based Feature Selection prediction model is formulated as given below.

$y_i = \beta_0 + \sum_{j=1}^m \beta_j + PD_j^i$  (7) From the above equation (7), ‘ $PD_j^i$ ’ forms the independent pre-processed air quality data, and ‘ $\beta$ ’ represents the regression coefficient and expected value of response (i.e., optimal and relevant feature selection) with the predictors (i.e., pre-processed air quality data) based on linear regression. ‘ $y_i$ ’ respectively. The optimum and relevant features are selected to minimize the sum of mean squared loss. This is mathematically formulated as given below.

$$\beta = \operatorname{argmin} L(Dis, \beta) = \operatorname{argmin} \sum_{i=1}^n (\beta \cdot PD_i - y_i)^2 \quad (8)$$



**Fig. 3 Flow of Linear Regression and Correlation-based Feature Selection**

From the above equation (8), the sum of mean squared loss ‘*L*’ is estimated based on the distance ‘*D*’ between two pre-processed data ‘*PD*’ via regression coefficients ‘ $\beta$ ’ taken for selecting relevant feature. Finally, the optimum relevant features are selected using a gradient as given below.

$$\frac{\partial L(Dis, \beta)}{\partial \beta} = \frac{\partial (Y^T Y - Y^T PD \beta - \beta^T PD^T PD \beta)}{\partial \beta} \quad (9)$$

From the above equation (9), the optimum relevant features with minimum loss are obtained by the partial differentiating correlation between pre-processed data or features based on distance ‘ $\partial L(Dis, \beta)$ ’ for each regression coefficient ‘ $\partial \beta$ ’. The pseudocode representation of Gradient Linear Regression-based Feature Selection is given below.

As given in the bel algorithm, Gradient Linear Regression-based Feature Selection is designed to improve the accuracy and time involved in air quality monitoring.

### 2.3 Multiclass Support Vector IoT-based Air Pollution Forecast model

Finally, in this section, with the relevant features selected (i.e., PM2.5, PM10, SO2, NOx, NH3, CO, O3), the classification of air quality data for the respective relevant features is made. The classification here is made through Air Quality Index (AQI) provided in table 1 to monitor the same and take preventive measures for controlling the same. Our work's classification uses a Multiclass Support Vector IoT-based Air Pollution Forecast model. Figure 4, given below, shows the structure of the Multiclass Support Vector IoT-based Air Pollution Forecast model.

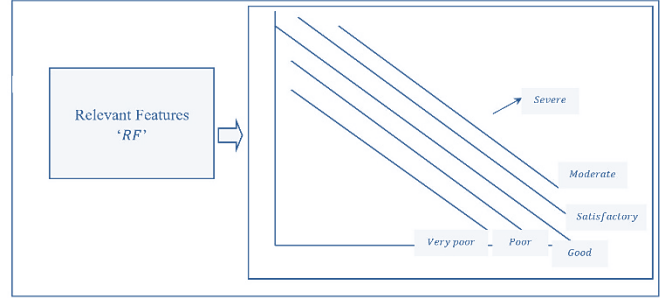


Fig. 4 Multiclass Support Vector IoT-based Air Pollution Forecast model

#### Algorithm.2 Gradient Linear Regression-based Feature Selection

<b>Input:</b> Dataset ‘DS’, Cloud Server ‘CS’, IoT Devices or Sensors ‘ $S = S_1, S_2, \dots, S_n$ ’, Features ‘ $F = F_1, F_2, \dots, F_n$ ’, Air Quality data ‘ $D = D_1, D_2, \dots, D_n$ ’
<b>Output:</b> Optimal and relevant feature selection
Step 1: <b>Initialize</b> pre-processed air quality data ‘PD’, regression coefficients ‘ $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ ’
Step 2: <b>Begin</b>
Step 3: <b>For</b> each pre-processed air quality data ‘PD’ with Cloud Server ‘CS’ and IoT Devices or Sensors ‘ $S = S_1, S_2, \dots, S_n$ ’
Step 4: Formulate linear regression function as given in equation (7)
Step 5: Evaluate the sum of mean squared loss as given in equation (8)
Step 6: Obtain optimum relevant features using the gradient function as given in equation (9)
Step 7: <b>Return</b> relevant features (RF)
Step 8: <b>End for</b>
Step 9: <b>End</b>

As shown in the above figure, given a dataset ‘DS’ with relevant features selected ‘ $RF = \{RF_1, RF_2, \dots, RF_n\}$ ’, ‘ $RF_i \in R^n$ ’ belong to ‘C’ different classes (i.e., ‘ $C = 6$ ’), forming a tuple. ‘ $(RF_i, Y_i)$ ’. In addition, let ‘ $Y_i \in \{0, 0 - 0.25, 0.25 - 0.50, 0.50 - 0.75, 0.75 - 1, > 1\}$ ’ representing 6 different classes, namely, good, satisfactory, moderate, poor, very poor and severe, respectively, is considered as its class labels. The recommendation confinements of the linear classifier are then mathematically formulated as given below.

$$W^T RF + B \quad (10)$$

From the above equation (10), ‘W’ denotes the weight vector and ‘B’ forms the bias vector, respectively. Several linear separators are present. However, the SVM design objective remains in determining a decision boundary that is maximally far away from any data point. This distance from the Decision boundary (i.e., AQI) to the closest data point (i.e., RF) determines the margin of the classifier. The air quality index for each sampled data is measured as given below.

$$AQI = Avg(PM2.5, PM10, SO2, NOx, NH3) + Max(CO, O3) \quad (11)$$

#### Algorithm. 3 Multiclass Support Vector IoT-based Air Pollution Forecast

<b>Input:</b> Dataset ‘DS’, Cloud Server ‘CS’, IoT Devices or Sensors ‘ $S = S_1, S_2, \dots, S_n$ ’, Features ‘ $F = F_1, F_2, \dots, F_n$ ’, Air Quality data ‘ $D = D_1, D_2, \dots, D_n$ ’
<b>Output:</b> Robust classification
Step 1: <b>Initialize</b> relevant features (RF)
Step 2: <b>Begin</b>
Step 3: <b>For</b> each relevant features ‘RF’ with Cloud Server ‘CS’ and IoT Devices or Sensors ‘ $S = S_1, S_2, \dots, S_n$ ’
Step 4: Formulate recommendation confinements of the linear classifier as in equation (10)
Step 5: Measure Air Quality Index as in equation (11)
Step 6: <b>If</b> ‘ $W^T RF_i(AQI) + b = 0$ ’
Step 7: <b>Then</b> AQI_Bucket ‘ $AQI_B \rightarrow Good$ ’
Step 8: <b>End if</b>
Step 9: <b>If</b> ‘ $W^T RF_i + b(AQI) > 0$ and $W^T RF_i + b(AQI) < 0.50$ ’
Step 10: <b>Then</b> AQI_Bucket ‘ $AQI_B \rightarrow Satisfactory$ ’
Step 11: <b>End if</b>
Step 12: <b>If</b> ‘ $W^T RF_i + b(AQI) > 0.50$ and $W^T RF_i + b(AQI) < 1$ ’
Step 13: <b>Then</b> AQI_Bucket ‘ $AQI_B \rightarrow Moderate$ ’
Step 14: <b>End if</b>
Step 15: <b>If</b> ‘ $W^T RF_i + b(AQI) > 0$ and $W^T RF_i + b(AQI) < -0.50$ ’
Step 16: <b>Then</b> AQI_Bucket ‘ $AQI_B \rightarrow Poor$ ’
Step 17: <b>End if</b>
Step 18: <b>If</b> ‘ $W^T RF_i + b(AQI) > -0.50$ and $W^T RF_i + b(AQI) < -1$ ’
Step 19: <b>Then</b> AQI_Bucket ‘ $AQI_B \rightarrow Very Poor$ ’
Step 20: <b>End if</b>
Step 21: <b>If</b> ‘ $W^T RF_i + b(AQI) > 1$ ’
Step 22: <b>Then</b> AQI_Bucket ‘ $AQI_B \rightarrow Severe$ ’
Step 23: <b>End if</b>
Step 24: <b>End for</b>
Step 25: <b>End</b>

From the above equation (11), the air quality index value ‘AQI’ is measured based on the average values of PM2.5, PM10, SO2, NOx, NH3 and the maximum values of CO and O3, respectively. These points are referred to as the support vectors. Finally, the resultant of AQI\_Bucket is estimated to

measure the air quality so that control measures can be made according to the pollutants. The pseudocode representation of the Multiclass Support Vector IoT-based Air Pollution Forecast is given below.

As given in the above Multiclass Support Vector IoT-based Air Pollution Forecast algorithm, the objective is to forecast air pollution with minimum error.

### 3. Experimental Setup

In this section, the performance of air pollution monitoring and control using IoT in a cloud computing environment method called Linear Regression and Multiclass Support Vector (LR-MSV) is performed using the Java interfaces and CloudSim simulator. To measure the LR-MSV method, air quality data in India via <https://www.kaggle.com/rohanrao/air-quality-data-in-india> is used. First, dataset details are provided. Following this, experiments were conducted on factors such as air pollution forecasting time, air pollution forecasting accuracy and error rate concerning different air quality samples. A fair comparison is made with the existing methods, Integrated Multiple Directed Attention and Variational Auto Encoder (VAE) (IMD-VAE) [1] and Bidirectional Recurrent Neural Network [bidirectional RNN] [2] for a simulation of 10 runs.

#### 3.1. Dataset Details

The Air Quality in India dataset contains air quality data and AQI (Air Quality Index) measured on both hourly and daily basis of several stations across multiple cities in India

Table 1 Air Quality in India dataset description

S. No	Features	Description
1	City	City name
2	Date	Date of occurrence
3	PM 2.5	Particulate Matter 2.5
4	PM 10	Particulate Matter 10
5	NO	Nitric Oxide
6	NO2	Nitric dioxide
7	NOx	Any nitric x-oxide
8	NH3	Ammonia
9	CO	Carbon monoxide
10	SO2	Sulphur dioxide
11	O3	Ozone
12	C6H6	Benzene
13	C7H8	Toluene
14	C8H10	Xylene

#### 3.2. Performance Analysis of Air Pollution Forecasting Time

Air pollution forecasting refers to applying science and technology to predict air pollution composition in the atmosphere for any location taken into consideration along with its respective time. Mainstream pollution forecasting

methods tend to utilize air quality index to indicate pollution levels, and also, in our work, it has been employed using a multiclass support vector. The air pollution forecasting time is mathematically formulated as given below.

$$APF_{time} = \sum_{i=1}^n D_i * Time [W^T RF + B] \quad (12)$$

From the above equation (12), the air pollution forecasting time ‘ $APF_{time}$ ’ is measured based on the air quality sample data involved in the simulation process. ‘ $D_i *$ ’ and the time consumed in the forecasting process via multiclass support vectors ‘ $Time [W^T RF + B]$ ’. It is measured in terms of milliseconds (ms). Table 2 below shows the performance analysis of the proposed air pollution monitoring and controlling method, LR-MSV, for air pollution forecasting time. The proposed method is applied to the air quality data obtained from different sensors and stored in the cloud server. In this work, the simulation results of the proposed air pollution monitoring and controlling method are compared with other state-of-the-art methods, IMD-VAE [1] and bidirectional RNN [2], which used similar air quality data in the India dataset.

Table 2. Tabulation for air pollution forecasting time

Air quality data	Air pollution forecasting time (ms)		
	LR-MSV	IMD-VAE	bidirectional RNN
2000	960	1100	1300
4000	1025	1255	1635
6000	1085	1315	1855
8000	1235	1535	1925
10000	1415	1895	2235
12000	1635	2135	2455
14000	1915	2325	2835
16000	2245	2635	3015
18000	2585	3025	3455
20000	3025	3355	3825

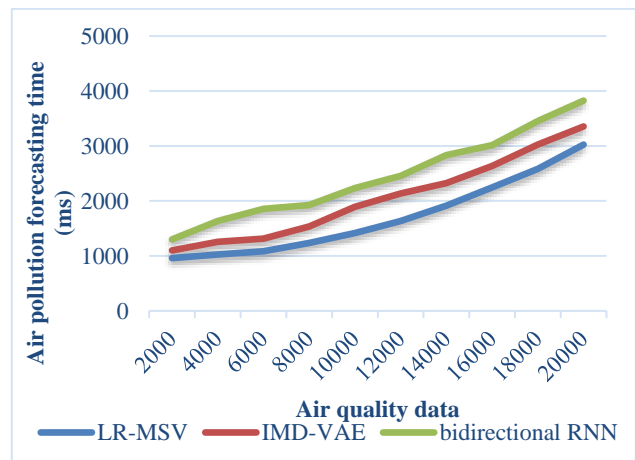


Fig. 5 Graphical representation of air pollution forecasting time

Figure 5 above shows the performance metric of air pollution forecasting time for varying air quality data obtained

from distinct sensors, each obtaining different air quality data in the range of 2000 to 20000 collected at different time intervals. From the figure, it is inferred that the air pollution forecasting time increases with the increase in the number of air quality data. This is because increasing the number of air quality data causes an increase in the number of data involved in pre-processing, which also increases the air pollution forecasting time. However, With ‘2000’ air quality data involved in simulation and the time consumed in air pollution forecasting for single data being ‘0.48ms’ using LR-MSV, the overall air pollution forecasting time was observed to be ‘960ms’, time consumed in air pollution forecasting being ‘0.55ms’ using [1] and ‘0.65ms’ using [2], the overall air pollution forecasting time was observed to be ‘1100ms’ and ‘1300ms’. From the result, the air pollution forecasting time via pre-processing was found to be minimum using LR-MSV when compared to [1] and [2]. The reason behind the improvement was due to the application of the Time and Frequency-based Sliding Window Pre-processing algorithm.

**3.3. Performance Analysis of Air Pollution Forecasting Accuracy**

The second parameter of significance for air pollution monitoring and control is air pollution forecasting accuracy. This parameter is highly significant because the more accurate the air pollution forecasting being made earlier the control measure done and therefore preventing hazardous disease. The mathematical formulation for air pollution forecasting accuracy is given below.

$$APF_{acc} = \sum_{i=1}^n \frac{D_{FAcc}}{D_i} * 100 \quad (13)$$

Table 3. Tabulation for air pollution forecasting accuracy

Air quality data	Air pollution forecasting accuracy (%)		
	LR-MSV	IMD-VAE	bidirectional RNN
2000	92.25	85.75	83.75
4000	91.45	83.15	81.45
6000	91	81	78.15
8000	90.25	78.35	75.35
10000	88.35	78	71
12000	85.25	77.35	70.25
14000	83.15	75.25	68.35
16000	83	75	65
18000	82.55	73.15	63.15
20000	80	73	60

From the above equation (13), the air pollution forecasting accuracy. ‘ $APF_{acc}$ ’ is measured based on the air quality sample data considered for simulation. ‘ $D_i$ ’ and the air quality data was forecasted accurately. ‘ $D_{FAcc}$ ’. It is measured in terms of percentage (%). Table 3 compares the proposed air pollution monitoring and control system with state of arts IMD-VAE [1] and bidirectional RNN [2]. The traditional air

pollution monitoring and control methods utilized air quality data, and the results are tabulated in Table 3. From Table 3, it is very clear that the air pollution monitoring system achieves maximum accuracy.

Second, figure 6 given above demonstrates the air pollution forecasting accuracy for 20000 varying numbers of air quality data. From the figure, it is inferred that the accuracy of air pollution forecasting decreases with the increase in air quality data. This is because a small portion of the presence of noise or artefacts is said to be retained during the pre-processing stage, and accuracy variation is also said to be identified. However, simulations conducted with ‘2000’ air quality data ‘1845’ air quality data were correctly detected as it is using LR-MSV, whereas ‘1715’ and ‘1675’ were correctly detected using [1] and [2]. From this analysis, the air pollution forecasting accuracy was found to be ‘92.25%’, ‘85.75%’ and ‘83.75%’ using LR-MSV, [1] and [2] respectively. From this result, it is inferred that the air pollution forecasting accuracy using the LR-MSV method is said to be comparatively better than [1] and [2]. The reason behind the improvement is due to the incorporation of the Gradient Linear Regression-based Feature Selection algorithm. By applying this algorithm, air quality data forecasted accurately that form the basis for early and precise air pollution monitoring and control is said to be obtained via the Gradient Linear Regression function for each pre-processed air quality data. Then, the sum of the mean squared loss for each feature with the adjacent is made for selecting the best feature for air pollution monitoring. With this, the air pollution forecasting accuracy using the LR-MSV method is said to be improved by 11% compared to [1] and 22% compared to [2].

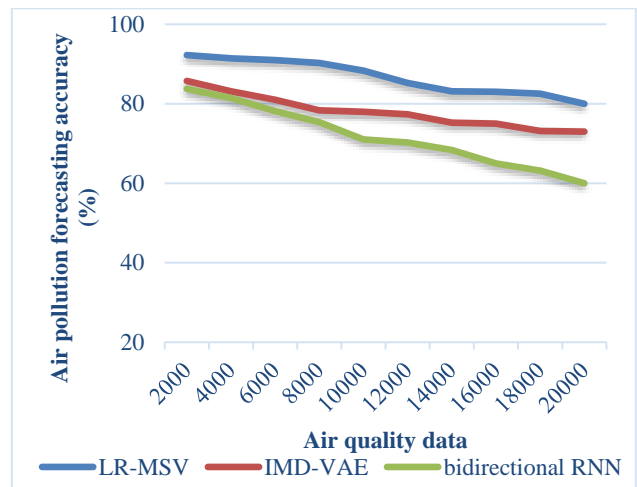


Fig. 6 Graphical representation of air pollution forecasting accuracy

**3.4. Performance Analysis of Error Rate**

Finally, the metric of significance that has to be measured for air pollution monitoring and control is the error rate. The error rate is mathematically formulated as given below.

$$ER = \sum_{i=1}^n \frac{D_{FWrongly}}{D_i} * 100 \quad (14)$$

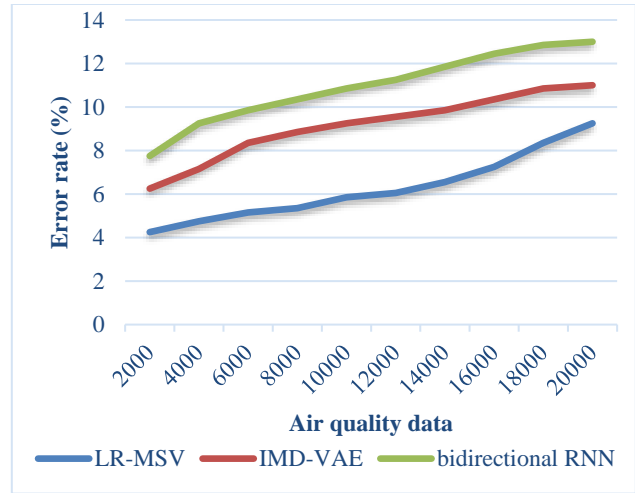
From the above equation (14), the error rate ‘ER’ is measured based on the air quality data samples involved in the simulation process. ‘ $D_i$ ’ and the air quality data was forecasted wrongly. ‘ $D_{FWrongly}$ ’. This is measured in terms of percentage (%). Table 4 compares the proposed air pollution monitoring and control error rate with other state-of-the-art methods [1] and [2], respectively.

**Table 4. Tabulation for error rate**

Air quality data	Error rate (%)		
	LR-MSV	IMD-VAE	bidirectional RNN
2000	4.25	6.25	7.75
4000	4.75	7.15	9.25
6000	5.15	8.35	9.85
8000	5.35	8.85	10.35
10000	5.85	9.25	10.85
12000	6.05	9.55	11.25
14000	6.55	9.85	11.85
16000	7.25	10.35	12.45
18000	8.35	10.85	12.85
20000	9.25	11	13

Finally, figure 7 given above shows the error rate for 20000 distinct air quality data obtained at different time instances. With error rate increased with the increasing air quality samples but was found to be comparatively reduced using LR-MSV upon comparison with [1] and [2]. However, simulations performed with 20000 samples observed that 85 samples were wrongly forecasted using LR-MSV, 125 samples were wrongly forecasted using [1], and 155 samples were wrongly forecasted using [2]. With this, the overall error rate was 4.25% using the LR-MSV method, 6.25% using [1] and 7.75% using [2]. The minimum error rate contribution using the LR-MSV method was due to the application of the Multiclass Support Vector IoT-based Air Pollution Forecast algorithm. By applying this algorithm, the air quality index was first measured and then supported vectors were obtained based on multiclass, therefore reducing the error rate using the

LR-MSV method by 32% compared to [1] and 43% compared to [2], respectively.



**Fig. 7 Graphical representation of error rate**

#### 4. Conclusion

The machine learning techniques serve as a refining service to confront air pollution forecast and control, progressively increasing the attention of researchers and academics. Inferences from previous research work specify that there is a requirement to design an effective method to provide accurate forecasting and control measures to be taken accordingly based on the air pollution via air quality index. Specifically, there is a requirement to address time and accuracy involving air pollution forecasting to prevent hazardous effects on humans. Hence, this work aims to address air pollution monitoring and control via machine learning technique to develop a Linear Regression and Multiclass Support Vector (LR-MSV) IoT-based Air Pollution Forecast method that provides forecasts in an accurate and timely manner with a minimum error rate. The experimental results show that the LR-MSV method can get better results in terms of air pollution forecasting accuracy, time and error rate on air quality data in the India dataset than others, which fully shows that applying it to machine learning can improve the forecasting accuracy and therefore paving the way for significant control.

#### References

[1] Abdelkader Dairi, Fouzi Harrou, Sofiane Khadraoui and Ying Sun, “Integrated Multiple Directed Attention-based Deep Learning for Improved Air Pollution Forecasting,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, 2021. Crossref, <https://doi.org/10.1109/TIM.2021.3091511>

[2] D. Saravanan and K. Santhosh Kumar, “Improving Air Pollution Detection Accuracy and Quality Monitoring Based on Bidirectional RNN and the Internet of Things,” *Materials Today: Proceedings, Elsevier*, 2021. Crossref, <https://doi.org/10.1016/j.matpr.2021.04.239>

[3] Mauro Castelli, Fabiana Martins Clemente, Ales Popovic, Sara Silva and Leonardo Vanneschi, “A Machine Learning Approach to Predict Air Quality in California,” *Complexity*, Wiley, 2020. Crossref, <https://doi.org/10.1155/2020/8049504>

- [4] WenTsai Sung and SungJung Hsiao, "Building an Indoor Air Quality Monitoring system Based on the Architecture of the Internet of Things," *EURASIP Journal on Wireless Communications and Networking*, Springer, 2021. Crossref, <https://doi.org/10.1186/s13638-021-02030-1>
- [5] Koel Datta Purkayastha, Ritesh Kishore Mishra, Arunava Shil and Sambhu Nath Pradhan, "IoT Based Design of Air Quality Monitoring System Web Server for Android Platform," *Wireless Personal Communications*, Springer, vol. 118, pp. 2921–2940, 2021. Crossref, <https://doi.org/10.1007/s11277-021-08162-3>
- [6] Ahmed Samy Moursi, Nawal El-Fishawy, Soufiene Djahel and Marwa Ahmed Shouman, "An IoT enabled system for Enhanced Air Quality Monitoring and Prediction on the Edge," *Complex & Intelligent Systems*, Springer, 7, pp. 2923-2947, 2021. Crossref, <https://doi.org/10.1007/s40747-021-00476-w>
- [7] D. Zhang and S. S. Woo, "Real Time Localized Air Quality Monitoring and Prediction Through Mobile and Fixed IoT Sensing Network," *IEEE Access*, vol. 8, pp. 89584-89594, 2020, Crossref, <https://doi.org/10.1109/ACCESS.2020.2993547>.
- [8] D. Hepsiba, L.M. Varalakshmi, M. Suresh Kumar, Jayasudha Subburaj, P. Vijayakumar and S. Sheeba Rani, "Automatic Pollution Sensing and Control for Vehicles Using IoT technology," *Materials Today: Proceedings*, Elsevier, vol. 45, pp. 3019-3021, 2021. Crossref, <https://doi.org/10.1016/j.matpr.2020.12.007>
- [9] Raquel Espinosa, José Palma, Fernando Jiménez, Joanna Kamińska, Guido Sciacivico and Estrella Lucena-Sánchez, "A Time Series Forecasting Based Multi-Criteria Methodology for Air quality Prediction," *Applied Soft Computing*, Elsevier, vol. 113, 2021. Crossref, <https://doi.org/10.1016/j.asoc.2021.107850>
- [10] Uzair Aslam Bhatti, Yuhuan Yan, Mingquan Zhou, Sajid Ali, Aamir Hussain, Huo Qingsong, Zhaoyuan Yu and Lingwang Yuan, "Time Series Analysis and Forecasting of Air Pollution Particulate Matter (PM<sub>2.5</sub>): An SARIMA and Factor Analysis Approach," *IEEE Access*, vol. 9, pp. 41019-41031, 2021, Crossref, <https://doi.org/ACCESS.2021.3060744>.
- [11] Prithijit Nath, Pratik Saha, Asiflqbal Middy and Sarbani Roy, "Long-Term Time-Series Pollution Forecast Using Statistical and Deep learning Methods," *Neural Computing and Applications*, Springer, 33, pp. 12551–12570, 2021. Crossref, <https://doi.org/10.1007/s00521-021-05901-2>
- [12] ShreyaKusrey, Avinash Rai and Dr. Vineeta (Nigam) Saxena, "Zigbee Based Air Pollution Monitoring and Control System using WSN," *SSRG International Journal of Electronics and Communication Engineering*, vol. 4, no. 6, pp. 7-11, 2017. Crossref, <https://doi.org/10.14445/23488549/IJECE-V4I6P103>
- [13] Mario Catalanao, Fabio Galatiotob, Margaret Bellb, Anil Namdeob and Angela S. Bergantinoc, "Improving the Prediction of Air Pollution Peak Episodes Generated by Urban Transport Networks," *Environmental Science & Policy*, Elsevier, vol. 60, pp. 69-83, 2016. Crossref, <https://doi.org/10.1016/j.envsci.2016.03.008>
- [14] Nabil Mohamed Eldakhly, MagdyAboul-Ela and AreegAbdalla, "A Novel Approach of Weighted Support Vector Machine with Applied Chance Theory for Forecasting Air Pollution Phenomenon in Egypt," *International Journal of Computational Intelligence and Applications*, vol. 17, no. 1, 2018. Crossref, <https://doi.org/10.1142/S1469026818500013>
- [15] Amina Nazif, Nurullzma Mohammed, Amirhossein Malakahmad and Motasem S. Abualqumboz, "Regression and Multivariate Models for Predicting particulate Matter Concentration Level," *Environmental Science and Pollution Research*, Springer, 25, pp. 283–289, 2018. Crossref, <https://doi.org/10.1007/s11356-017-0407-2>
- [16] Murtaza Ali Khan, "Detection and Classification of Plant Diseases Using Image Processing and Multiclass Support Vector Machine," *International Journal of Computer Trends and Technology*, vol. 68, no. 4, pp. 5-11, 2020. Crossref, <https://doi.org/10.14445/22312803/IJCTT-V68I4P102>
- [17] Anikender, Kumar and Pramila Goyal, "Forecasting of Air Quality in Delhi Using Principal Component Regression Technique," *Atmospheric Pollution Research*, vol. 2, no. 4, pp. 436-444, 2011. Crossref, <https://doi.org/10.5094/APR.2011.050>
- [18] R.Amulraju, D. Ashok Kumar and T.Vithyaa, "A Comparative Study on Air Quality Analysis Through DNN by SVM, K - Means and Naive Bayes Algorithms," *International Journal of Computer Trends and Technology*, vol. 67, no. 11, pp. 42-48, 2019. Crossref, <https://doi.org/10.14445/22312803/IJCTT-V67I11P107>
- [19] Tong Liu, Alexis K. H. Lau, Kai Sandbrink, and Jimmy C. H. Fung, "Time Series Forecasting of Air Quality Based On Regional Numerical Modeling in Hong Kong," *Journal of Geophysical Research: Atmospheres*, vol. 123, no. 3, 2018. Crossref, <https://doi.org/10.1002/2017JD028052>
- [20] L. Jennifer, Moutinho, Donghai Liang, Rachel Golan, Stefanie E. Sarnat, Rodney Weber, Jeremy A. Sarnat and Armistead G. Russell, "Near-Road Vehicle Emissions Air Quality Monitoring for Exposure Modeling," *Atmospheric Environment*, Elsevier, vol. 224, 2020. Crossref, <https://doi.org/10.1016/j.atmosenv.2020.117318>
- [21] Isura Sumeda Priyadarshana Nagahage, Ekanayaka Achchillage Ayesha Dilrukshi Nagahage, Takeshi Fujino, "Assessment of the Applicability of a Low-Cost Sensor-Based Methane Monitoring System for Continuous Multi-Channel Sampling," *Environmental Monitoring and Assessment*, Springer 193, pp. 509, 2021. Crossref, <https://doi.org/10.1007/s10661-021-09290-w>

- [22] R. Smit, P. Kingstona, D.W. Nealea, M.K. Browna, B. Verrana and T. Nolana, "Monitoring On-Road Air Quality and Measuring Vehicle Emissions with Remote Sensing in an Urban Area," *Atmospheric Environment, Elsevier*, vol. 218, 2019. Crossref, <https://doi.org/10.1016/j.atmosenv.2019.116978>
- [23] A.Aarathi, P.Gayathri, N.R.Gomathi, S.Kalaiselvi and Dr.V.Gomathi, "Air Quality Prediction Through Regression Model," *International Journal of Scientific & Technology Research*, vol. 9, no. 3, 2020.

Original Article

# Novel Regression and Least Square Support Vector Machine Learning Technique for Air Pollution Forecasting

M. Dhanalakshmi<sup>1</sup>, V. Radha<sup>2</sup>

<sup>1,2</sup>Department of Computer Science, Avinashilingam Institute of Home Science and Higher Education for Women, Coimbatore, India.

<sup>1</sup>Corresponding Author : [dhanalaxmi2289@gmail.com](mailto:dhanalaxmi2289@gmail.com)

Received: 23 November 2022

Revised: 01 April 2023

Accepted: 17 April 2023

Published: 25 April 2023

**Abstract** - Air pollution is the origination of particulate matter, chemicals, or biological substances that brings pain to either humans or other living creatures or instigates discomfort to the natural habitat and the airspace. Hence, air pollution remains one of the paramount environmental issues as far as metropolitan cities are concerned. Several air pollution benchmarks are even said to have a negative influence on human health. Also, improper detection of air pollution benchmarks results in severe complications for humans and living creatures. To address this aspect, a novel technique called, Discretized Regression and Least Square Support Vector (DR-LSSV) based air pollution forecasting is proposed. The results indicate that the proposed DR-LSSV Technique can efficiently enhance air pollution forecasting performance and outperforms the conventional machine learning methods in terms of air pollution forecasting accuracy, air pollution forecasting time, and false positive rate.

**Keywords** - Air pollution monitoring, Discretized hartley transformation, Constrained maximum likelihood, Linear regression, SVM, Air pollution forecasting, Novel machine learning algorithms.

## 1. Introduction

Air pollution presents a significant health problem as far as urban metropolises are concerned. Though air pollution monitoring and forecasting accurately and precisely are found to be tremendously critical, prevailing data-driven methods have so far entirely acquired the complicated interactions between spatial and temporal aspects of air pollution. Moreover, uneasiness for the environment, health, and welfare has fascinated substantial global awareness owing to the new environmental confronting that menace the planet.

A method called Deep-AIR employing CNN and LSTM framework was proposed by Qi Zhang et al. [1] to fill the gap in ensuring fine-grained city-wide air pollution estimation by means of domain-specific features that, in turn, captured spatio-temporal features. Also, a 1 to 1 convolution layer was designed with the purpose of improving the learning of temporal and spatial interaction. As a result, the forecasting accuracy for air pollution was improved.

Variational Auto Encoder (VAE) based on the innovative Integrated Multiple Direct Attention Deep Learning architecture (IMDA) [VAE-IMDA] was proposed by Abdelkader et al. [2], taking into consideration the conventional VAE and attention mechanism to forecast

distinct air pollutants in a computationally efficient and accurate manner. Also, the temporal dependencies between nonlinear approximations concentrating on the relevant feature extraction were ensured.

To this extent, this paper is proposed for a research study on applying Discretized Regression and Least Square Support Vector (DR-LSSV) based air pollution monitoring and control for IoT networks. In this study, we proposed a system for predicting air quality by advanced machine learning model - hybrid Discretized Regression and Least Square Support Vector air pollution forecasting. Two baseline models were also built for comparison with our proposed method.

## 2. Literature Review

Up till now, it remains a major issue in acquiring optimal and accurate air pollution estimation with high accuracy and low false positive rate simultaneously. Numerous missing values are said to exist in both temporal and spatial dimensions. This, in turn, has resulted in severely undermining the performance of optimization-based air pollution methods.

A two-layer model prediction method based on Long Short-Term Memory Neural Network and Gated Recurrent



Unit called LSTM&GRU was proposed by Baowei et al. [3]. A double-layer Recurrent Neural Network method was designed with the objective of predicting the PM2.5 value. With this two-layer model, a better prediction rate was achieved. A novel deep learning-based air quality forecasting by learning spatial-temporal correlation features and multivariate air quality-related interdependence time series data employing hybrid deep learning architecture was proposed by shengdong et al. [4]. Satisfied accuracy was said to be achieved via a hybrid learning model. Despite accuracy being attained, the precision was not focused. To concentrate on this issue, a hybrid CNN-LSTM model was proposed by Shreya K et al.[5] that not only ensured accurate prediction via extraction of spatiotemporal features using CNN but also improved the predictive performance via LSTM-based deep learning.

The ceaseless precipitation of global urbanization and industrialization has led the way towards several issues. One of the important environmental issues is air quality, persuaded by the deployment of urbanization and industrialization. In Yuting et al.[7], explainable deep learning model was introduced with the objective of improving the prediction accuracy of air quality. In Qing et al.[8], yet another short-term forecasting model employing convolutional-based bidirectional gated recurrent unit (CBGRU) was proposed therefore minimizing the error involved during forecasting. A novel method to analyze deep air quality by utilizing convolutional neural and long short-term memory was proposed by Ekta et al.[9].

The prevailing and recent methods, however, only concentrated on giving detail about their causes and temporal associations. However, a holistic view of the pure forecasting performances was not analyzed. In Raquel et al. [11], a methodology was designed based on two criteria, namely, exactness and robustness, to make an elaborate comparison of distinct pollutant forecasting methods and their characteristics. With this high influence on accuracy was said to be arrived at.

Among numerous air pollutants, Particulate Matter of diameter less than 2.5 $\mu$ m is one of the most serious health issues. It results in different types of illnesses in the respiratory tract and cardiovascular diseases. Therefore, it becomes highly necessary to predict PM2.5 concentrations accurately so that citizens can be circumvented the dangerous influence of air pollution in the initial stage. In abdellatif et al. [12], a deep learning solution was designed with the purpose of predicting the PM2.5 hourly forecast in Beijing, China. This work designed a hybrid model combining CNN-LSTM where spatial-temporal features were extracted by integrated historical pollutants, specifically, PM2.5 concentration in the neighboring stations. A case study of air pollution prediction using machine learning was investigated in detail by kumar et al. [13].

In Azim et al. [15], a novel new hybrid intelligent method was designed based on long short-term memory (LSTM) and a multi-verse optimization algorithm (MVO). With this hybrid intelligent method, air pollution prediction and analysis were made by utilizing the data from Combined Cycle Power Plants (CCPP). Moreover, a long short-term memory model was also utilized with the purpose of forecasting the amount of NO2 and SO2 by CCPP. Also, the MVO algorithm was employed for optimizing LSTM parameters to minimize errors involved in forecasting.

With several features involved in predicting air quality data, a small deviation would cause a greater amount of error. To address this aspect, a method for predicting air quality based on multiple data features by fusing multiple machine learning models was presented by Ying et al. [16]. With this, not only accuracy was ensured but also minimized the loss involved during air pollution forecasting. In Yue-shan et al. [17], an Aggregated LSTM model (ALSTM) was proposed by integrating local air quality monitoring stations, i.e., the station in adjacent areas and obtained from external pollution sources were combined. Also, to enhance the prediction accuracy, three LSTM models were aggregated for early predictions based on external sources of pollution obtained from adjacent industrial air quality stations. As a result, the prediction accuracy was also improved to a greater extent.

Owing to the vigorous, dynamic phenomenon and several spatio-temporal factors influencing air pollution dispersion and involving uncertainty estimates makes, it is trustworthy. This, in turn, assists the decision-makers in taking proper actions concerning the pollution crisis. In Ichrak et al. [19], a multi-point deep learning method was proposed by considering convolutional long short-term memory (ConvLSTM) for the large arbitrary nature of air quality forecasting. As a result, even accuracy was arrived at even though with the huge climatic change.

As far as the meteorological forecast and air controlling are concerned, air quality prediction is contemplated as the paramount reference. However, with the occurrence of overfitting in prediction algorithms based on a single model, complexity is also said to be increased. To address this aspect, a prediction method based on integrated dual LSTM was proposed by Hongqian et al.[20]. With this, the precision of prediction data was said to be improved greatly. A recurrent neural network with LSTM was integrated by Saba et al. [21] with the purpose of not only reducing the error but also improving forecasting accuracy to a greater extent.

However, another method to minimize overfitting employing an ensemble network (EN) that combines recurrent neural network (RNN), LSTM network and gated recurrent unit (GRU) network was designed by Canyang et

al. [22]. With this ensemble type of networking, accuracy with the mean absolute error was improved greatly. In Jun et al. [23], urban ecological monitoring focused on deep learning to improve air pollution forecasting.[14] To identify nonlinear relationships between input and output variables, wavelet neural networks were integrated with meteorological conditions for air pollution forecasting by Qingchun et al. [24]. To address air pollution monitoring and control via machine learning technique[6], a Linear Regression and Multiclass Support Vector (LR-MSV) IoT-based Air Pollution Forecast method provide forecasts in an accurate and timely manner with minimum error rate was designed in Dhanalakshmi et al. [12].

In this paper, by a comparison of conventional machine learning and classical deep learning models, a novel air quality forecasting method, called Discretized Regression and Least Square Support Vector (DR-LSSV), is proposed. It is motivated to address error minimization and improve accuracy by utilizing the Discretized Hartley Transformation and performing feature selection using Constrained Maximum Likelihood Linear Regression. Finally, the proposed DR-LSSV method can classify air quality-related time series data by means of the air quality index and the selected features via the Concordance Correlative function for different weather conditions and different traffic states both on an hourly and daily basis.

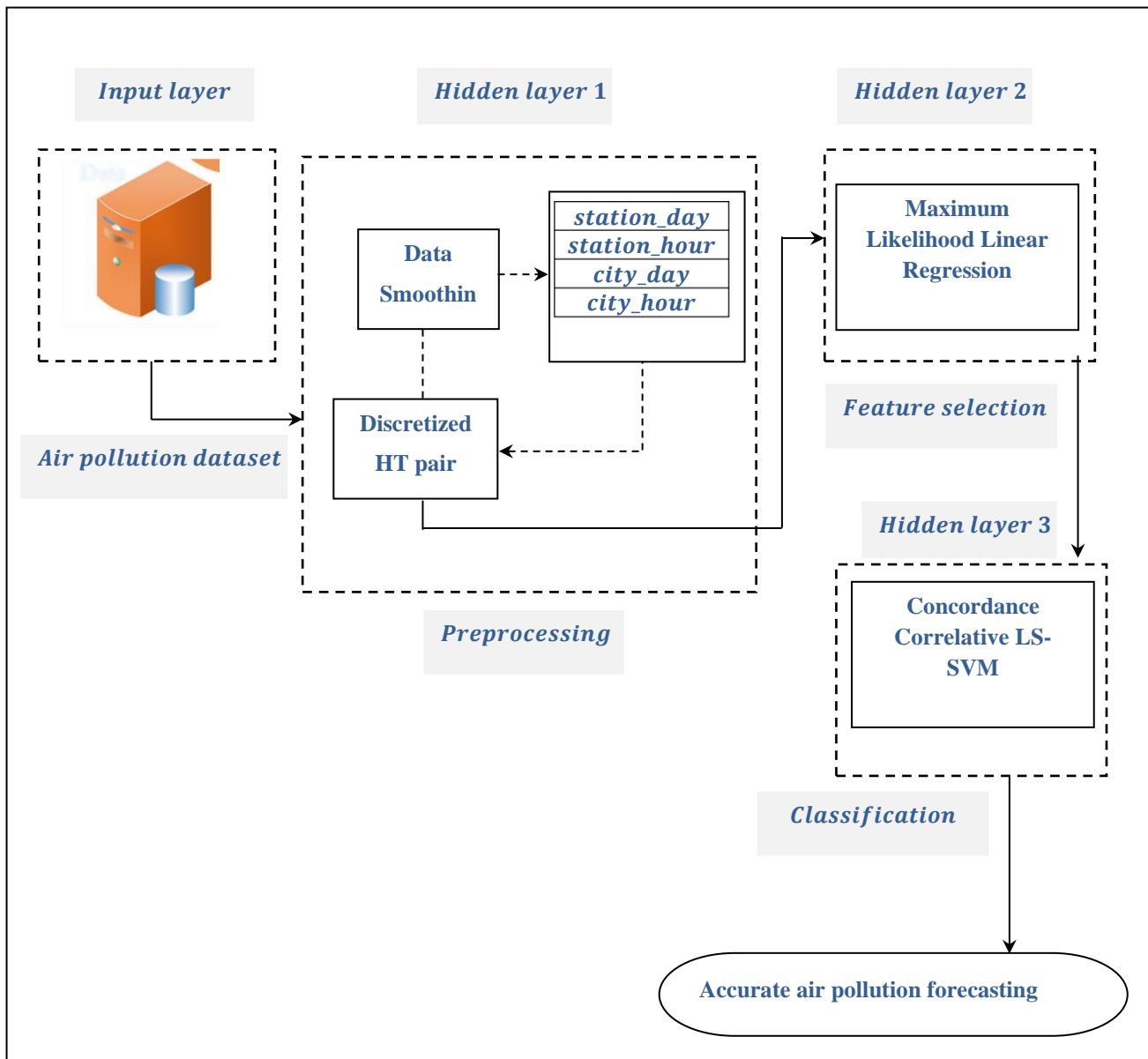


Fig. 1 Architecture of Discretized Regression and Least Square Support Vector

### 3. Methodology

Air pollution forecasting refers to the application of science and technology with the objective of predicting the air pollution composition present in the atmosphere for a specific location and time. Over the past decades, air pollution has been considered the world’s considerable issue. It brings about respiratory-related issues, cardiovascular concerns, lung diseases, and therefore resulting in mental-related affairs and provoking prevailing health state of affairs. Hence, minimizing the making of people well informed of these issues emanated by air pollution becomes indispensable. In this work, a novel method called Discretized Regression and Least Square Support Vector (DR-LSSV) is proposed in this work to increase air pollution forecasting accuracy with minimizing error. Figure 1, given below, shows the overall architecture of the DR-LSSV method.

As shown in the above figure, the proposed DR-LSSV method consists of three different processes, namely preprocessing, feature selection, and classification, with multiple layers involved, such as one input layer, three hidden layers, and one output layer. Air quality data in our work are obtained by means of IoT devices. The collected air

quality data are provided as input to the input layer. Air quality data obtained as input is provided in the first hidden layer, where preprocessing is carried out with the aid of Discretized Hartley Transformation function to remove the noise present in the raw air quality data.

Following the second hidden layer, the feature selection process is carried out using the Constrained Maximum Likelihood Linear Regression function that, in turn, selects significant features. Finally, the classification is performed in the third hidden layer by applying Concordance Correlative Least Square Support Vector to obtain final classification results by analyzing testing and training data. Based on the obtained results in the output layer, accurate air pollution forecasting is achieved with minimum error.

#### 3.1. Data Collection

This research work gathers information from Air Quality Data in India for the period between 2015 and 2020 that consists of Air Quality Index (AQI) at the hourly and daily levels of numerous stations in multiple cities in India. The AQI calculation utilizes 7 measures, namely, PM2.5, PM10, SO2, NO<sub>x</sub>, NH<sub>3</sub>, CO and O<sub>3</sub>.

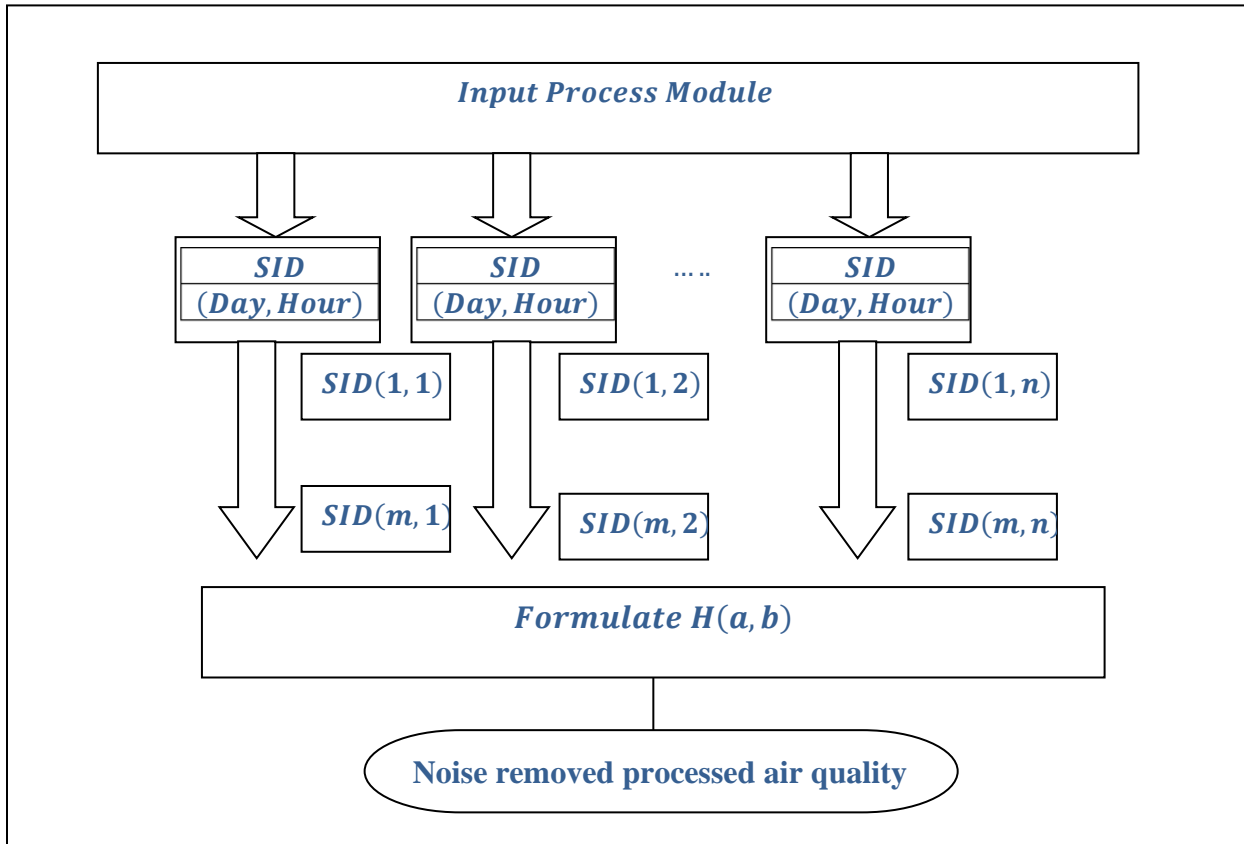


Fig. 2 Structure of Discretized Hartley Transformation-based Preprocessing model

### 3.2. Discretized Hartley Transformation-based Preprocessing model

Preprocessing is very important to environmental data because it influences the results of any data investigation method and, therefore, the performance of any data mining algorithm towards problem investigation and parameter forecasting. Preprocessing is very important to environmental data because it influences the results of any data investigation method and, therefore, the performance of any data mining algorithm towards problem investigation and parameter forecasting. Preprocessing is essential as far as environmental data is concerned owing to the reason that it has a great impact on the data investigation method and, therefore, the performance of any machine learning technique towards air pollution forecasting. It is frequent in air quality data to come across noise and extract real trends and patterns. With the objective of obtaining comparable results in the proposed method, the rule for replacing using Discretized Hartley Transformation-based Preprocessing model is employed in our work. The Discretized Hartley Transformation-based Preprocessing model proposed in our work transforms real inputs into real outputs. Figure 2, given below, shows the structure of Discretized Hartley Transformation-based Preprocessing model.

As given in the above figure, to extract real trends and patterns and eliminate noise, various cities in India are obtained in the first hidden layer with the station details obtained from the corresponding ID. Then, outliers or noise are removed by means of transforming real inputs to real outputs via the Discretized Hartley Transformation function.

Let us consider the input air quality data ‘ $D = D_1, D_2, \dots, D_n$ ’ obtained from sensors ‘ $S = S_1, S_2, \dots, S_n$ ’ be sent into the first hidden layer. The input air quality data here is stored in the form of vector-matrix separately for the corresponding station according to ‘ $Day$ ’ and hour ‘ $Hour$ ’ as given below.

$$SID[Day] = Day [D_1, D_2, \dots, D_n] \quad (1)$$

$$SID[Hour] = Hour [D_1, D_2, \dots, D_n] \quad (2)$$

Then, from the above formulations from equations (1) and (2), two dimensional DHT pair is formulated as given below.

$$H(a, b) = \sum_{Day=1}^m \sum_{Hour=1}^n SID (Day, Hour) cas \left[ 2\pi \left( \frac{aDay}{P} + \frac{bHour}{Q} \right) \right] \quad (3)$$

$$SID (Day, Hour) = \sum_{Day=1}^m \sum_{Hour=1}^n H(a, b) cas \left[ 2\pi \left( \frac{aDay}{P} + \frac{bHour}{Q} \right) \right] \quad (4)$$

From the above equations (3) and (4), ‘ $SID (Day, Hour)$ ’ denotes the DHT pair averaged over the hours of the day, and the DHT pair averaged and ‘ $H(a, b)$ ’ represents the Hartley spectrum coefficient, Also ‘ $P$ ’ and ‘ $Q$ ’ represents the rows and columns of the air quality data and ‘ $cas(\theta) = \cos(\theta) + \sin(\theta)$ ’ and is mathematically formulated as given below.

$$PD = H(a, b) = \sum_{Day=1}^m \sum_{Hour=1}^n DIS(Day, Hour) \left( Cos \left[ 2\pi \left( \frac{aDay}{P} + \frac{bHour}{Q} \right) \right] + Sin \left[ 2\pi \left( \frac{aDay}{P} + \frac{bHour}{Q} \right) \right] \right) \quad (5)$$

From the results obtained in the above equation (5), the real inputs are transformed into real outputs, therefore eliminating the noise from air quality data. The pseudo-code representation of Discretized Hartley Transformation-based Preprocessing is given below.

**Algorithm 1. Discretized Hartley Transformation-based Preprocessing**

**Input:** Dataset ‘ $DS$ ’, IoT Devices or Sensors ‘ $S = S_1, S_2, \dots, S_n$ ’, Features ‘ $F = F_1, F_2, \dots, F_n$ ’, Air Quality data ‘ $D = D_1, D_2, \dots, D_n$ ’

**Output:** processed air quality data ‘ $PD$ ’

- Step 1: **Initialize** rows and columns of the air quality data ‘ $P$ ’ and ‘ $Q$ ’
- Step 2: **Begin**
- Step 3: **For** each Dataset ‘ $DS$ ’ with Sensors ‘ $S$ ’ and air quality data ‘ $D = D_1, D_2, \dots, D_n$ ’
- Step 4: Obtain vector matrix separately for the corresponding station based on the day ‘ $Day$ ’ and hour ‘ $Hour$ ’ as in equations (1) and (2)
- Step 5: Formulate two dimensional HT pair as in equations (3) and (4)
- Step 6: Obtain discretized HT pair as in equation (5) to transform real inputs to real outputs
- Step 7: **Return** processed air quality data ‘ $PD$ ’
- Step 8: **End for**
- Step 9: **End**

As given in the above algorithm, with the objective of minimizing the error by eliminating the noise and extracting real trends and patterns, Discretized Hartley Transformation function is applied. With the raw air quality data obtained as input from the input layer, the first hidden layer performs the preprocessing by means of first separately forming the vector matrix. Next, the two-dimensional HT pair is structured based on day-wise and hourly air quality data. Finally, the transformation is performed by discretizing with the sine and cosine functions for the two-dimensional HT pair, therefore obtaining the processed air quality data.

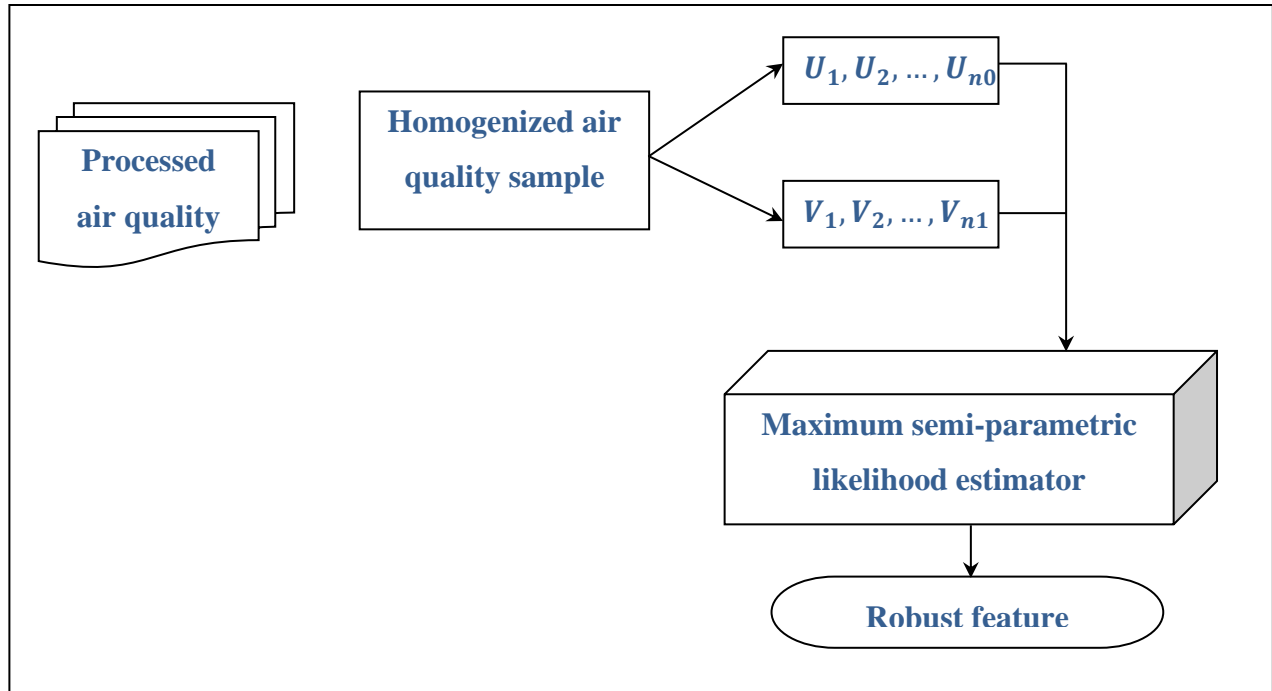


Fig. 3 Structure of Constrained Maximum Likelihood Linear Regression-based Feature Selection model

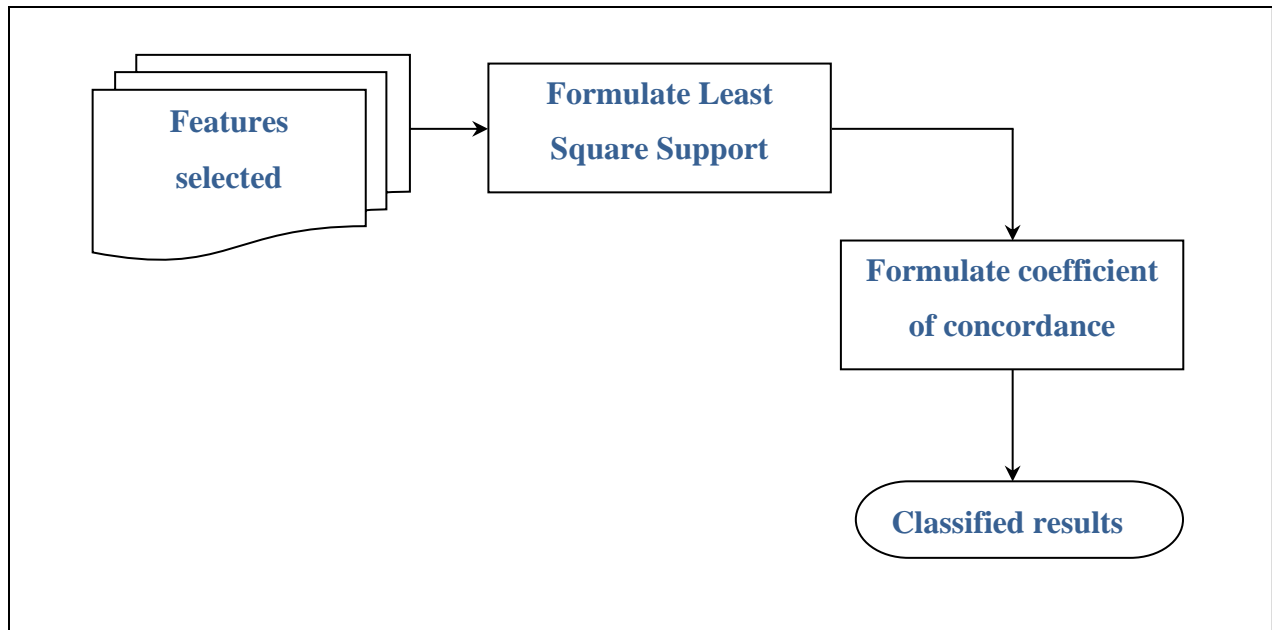


Fig. 4 Structure of Concordance Correlative Least Square Support Vector-based Classification

### 3.3. Constrained Maximum Likelihood Linear Regression-based Feature Selection model

For designing an efficient mechanism for tackling air pollution, it is essential to concentrate the endeavors on the pollutants most accountable for air pollution. Large numbers of processed air quality data features make the analysis a complicated task. Selection of the most essential and relevant features is the predominant step in air quality pollution monitoring and the decision-making process.

Several feature selection methods are present that aspire to minimize the number of dimensions in the processed air quality data to make the analysis simpler and more efficient. In this work, the feature selection process is carried out by employing Constrained Maximum Likelihood Linear Regression in the second hidden layer. Constrained Maximum Likelihood Linear Regression is one of the most popular models used for the selection of features by taking into consideration the correlative feature significance. The

feature significance is utilized in describing which features are most relevant or significant for air quality index (AQI) prediction by means of the maximum likelihood function. Figure 3 below shows the structure of the Constrained Maximum Likelihood Linear Regression-based Feature Selection model.

The above figure shows that the processed air quality data comprises redundant information. The constrained Maximum Likelihood Linear Regression criterion is utilized to select the maximum likelihood regressors that carry the maximal non-redundant information for air quality pollution and monitoring[10]. Initially, the processed air quality data provided as input to the second hidden layer is split into two homogenized vectors, one for station-wise air quality data on an hourly basis and another for station-wise quality data on a daily basis. Next, the two homogenized vectors are subjected to a Maximum semi-parametric likelihood estimator to select relevant and robust features.

Let  $(U_1, U_2, \dots, U_{n0})$  denote the arbitrary air quality sample data from  $Prob(u|V = 0)$  and  $(V_1, V_2, \dots, V_{n1})$  denote the arbitrary air quality sample data from  $Prob(u|V = 1)$  respectively. Also, let  $(T_1, T_2, \dots, T_n)$  represents the homogenized air quality sample data  $(U_1, U_2, \dots, U_{n0}; V_1, V_2, \dots, V_{n1})$ . Moreover, let  $(\alpha_0, \beta_0)$  denote the initial value of  $(\alpha, \beta)$ . Then, the logistic regression function for the homogenized air quality sample data with respect to the constrained  $(\alpha, \beta)$  is mathematically stated as given below.

$$Prob(V = 1|U = u) = \frac{\exp(\alpha + \beta^T T)}{1 + \exp(\alpha + \beta^T T)} \quad (6)$$

From the above equation (6),  $\alpha$  and  $\beta$ , represents the scale values. Then, the maximum semi-parametric likelihood estimator of  $(\alpha_0, \beta_0)$  is mathematically formulated as given below.

$$\frac{\partial l(\alpha, \beta)}{\partial \alpha} = \sum_{i=1}^n \frac{\exp(U_1, U_2, \dots, U_{n0})(\alpha + \beta^T T_i)}{\exp(\alpha + \beta^T T_i)} \quad (7)$$

$$\frac{\partial l(\alpha, \beta)}{\partial \beta} = \sum_{i=1}^n \frac{\exp(V_1, V_2, \dots, V_{n1})(\alpha + \beta^T T_i)}{\exp(\alpha + \beta^T T_i)} \quad (8)$$

From the results of the above two equations (7) and (8), with a maximum semi-parametric likelihood estimator applied to the regression results, redundant features are eliminated, and relevant features are selected for further processing.

$$FS = \frac{\partial l(\alpha, \beta)}{\partial \alpha} \cup \frac{\partial l(\alpha, \beta)}{\partial \beta} \quad (9)$$

From the above equation (9), finally, the significant features selected,  $FS$ , are obtained by combining the maximum semi-parametric likelihood estimator with respect to two constraints  $\partial \alpha$  and  $\partial \beta$ , respectively. The pseudo-code representation of Constrained Maximum Likelihood Linear Regression-based Feature Selection is given below.

**Algorithm 2. Constrained Maximum Likelihood Linear Regression-based Feature Selection**

<b>Input:</b> Dataset ‘ $DS$ ’, IoT Devices or Sensors ‘ $S = S_1, S_2, \dots, S_n$ ’, Features ‘ $F = F_1, F_2, \dots, F_n$ ’, Air Quality data ‘ $D = D_1, D_2, \dots, D_n$ ’
<b>Output:</b> Computationally efficient feature selection ‘ $FS$ ’
Step 1: <b>Initialize</b> processed air quality data ‘ $PD$ ’
Step 2: <b>Initialize</b> arbitrary air quality sample data ‘ $(U_1, U_2, \dots, U_{n0})$ ’ (i.e., processed station data at hourly basis)
Step 3: <b>Initialize</b> arbitrary air quality sample data ‘ $(V_1, V_2, \dots, V_{n0})$ ’ (i.e., processed station data at daily basis)
Step 4: <b>Begin</b>
Step 5: <b>For</b> each Dataset ‘ $DS$ ’ with Sensors ‘ $S$ ’ and processed air quality data ‘ $PD$ ’
Step 6: Evaluate the logistic regression function for the homogenized air quality sample data as in equation (6)
Step 7: Evaluate the maximum semi-parametric likelihood estimator as in equations (7) and (8)
Step 8: <b>Return</b> features selected ‘ $FS$ ’
Step 9: <b>End for</b>
Step 10: <b>End</b>

As given in the above algorithm, arbitrarily processed air quality is obtained to obtain computationally efficient features. Second, the logistic regression function is applied separately to make the analysis easier and simpler. Finally, the maximum semi-parametric likelihood estimator is modeled to select the significant and computationally efficient feature. [PM10, CO, O3]

**3.4. Concordance Correlative Least Square Support Vector-based Classification model**

Finally, with the selected significant features, classification has to be performed for accurate air pollution forecasting with the objective of minimizing the false positive rate. With the aid of the Air Quality Index (AQI), pollutant levels in the air can be determined. The range of AQI varies between ‘0 and 500’, with a higher number referring to lower air quality.

Numerous classification methods have been designed for air pollution forecasting; however, with AQI limits varying in nature, the inappropriate classification may cause adverse effects on human health. Hence, a model is required for precise air pollution forecasting with appropriate classification, therefore resulting in the improvement of the false positive rate. In this work, the Concordance Correlative Least Square Support Vector-based Classification model is applied in the third hidden layer to measure the nonlinear relationship between input variables and output variables with the least error.

With the measurement of the nonlinear relationship between input variables and output variables, the false positive rate or assessing the exposure of the target population to specific air pollutants analyzed is said to be precise. Moreover, applying the Concordance Correlative function establishes a nonlinear relationship between input variables and output variables, therefore obtaining the target population for specific air pollutants precisely. Figure 4 shows the structure of the Concordance Correlative Least Square Support Vector-based Classification model.

As shown in the figure 4, the Concordance Correlative Least Square Support Vector-based Classification model provides a linear equation solution with an enhancement in the objective function of conventional SVM. We employ 'FS<sub>k</sub>' as the three features selected with maximum semi-parametric likelihood estimator. Next, the selected features are subjected to the least square support vector. With the resultant values, the coefficient of concordance is applied to obtain the classified results. Let us consider the Least Square Support Vector-based Classification written as given below.

$$y(FS) = \omega^T \varphi(FS) + B \quad (10)$$

From the above equation (10), 'φ(FS)' forms the nonlinear mapping function for the selected features, with 'ω' forming the weight and 'B' the bias, respectively. The equation is then subjected to minimum correlation with respect to error 'Err' as given below.

$$\min(\omega, Err) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{k=1}^n Err_k^2 \quad (11)$$

$$\text{Subject to } y_k = \omega^T \varphi(FS_k) + B + Err_k \quad (12)$$

From the above equations (11) and (12), 'γ' forms the regularization parameter (i.e., setting index value) and 'Err' represents the error, respectively. The optimization equation for the coefficient of concordance using Kendall's Rank Correlation Coefficient is mathematically formulated as given below.

$$AQI = \tau = \frac{(N_{CP})\varphi(FS_k) - (N_{DP})\varphi(FS_k)}{\frac{n(n-1)}{2}} \quad (13)$$

From the above equation (13), any pair of observations '(Tr<sub>i</sub>, Ts<sub>i</sub>)' and '(Tr<sub>j</sub>, Ts<sub>j</sub>)' are said to be concordant if the sort order of '(Tr<sub>i</sub>, Tr<sub>j</sub>)' and '(Ts<sub>i</sub>, Ts<sub>j</sub>)' agrees, otherwise they are said to be discordant,  $\frac{n(n-1)}{2}$ , denotes the binomial coefficient to select two features from 'n' features. The denominator here represents the total number of pair combinations, hence coefficient ranging in between '-1 ≤ τ ≤ 1'. The pseudo code representation for Concordance Correlative Least Square Support Vector-based Classification is given below.

**Algorithm 3 Concordance Correlative Least Square Support Vector-based Classification**

Input: Dataset 'DS', IoT Devices or Sensors 'S = S<sub>1</sub>, S<sub>2</sub>, ..., S<sub>n</sub>', Features 'F = F<sub>1</sub>, F<sub>2</sub>, ..., F<sub>n</sub>', Air Quality data 'D = D<sub>1</sub>, D<sub>2</sub>, ..., D<sub>n</sub>'

Output:

Step 1: **Initialize** processed air quality data 'PD', features selected 'FS'

Step 2: **Begin**

Step 3: **For** each Dataset 'DS' with Sensors 'S', processed air quality data 'PD' and features selected 'FS'

Step 4: Formulate Least Square Support Vector-based Classification as in equation (10)

Step 5: Evaluate minimum correlation with respect to error as in equations (11) and (12)

Step 6: Measure Kendalls Rank Correlation Coefficient as in equation (13)

Step 7: **If** 'τ lies between - 1 and - 0.5'

Step 8: **Then** air quality is very poor

Step 9: **End if**

Step 10: **If** 'τ lies between - 5 and 0'

Step 11: **Then** air quality is poor

Step 12: **End if**

Step 13: **If** 'τ lies equals 0'

Step 14: **Then** air quality is good

Step 15: **End if**

Step 16: **If** 'τ lies between 0 and + 0.5'

Step 17: **Then** air quality is satisfactory

Step 18: **End if**

Step 19: **If** 'τ lies between + 0.5 and 1'

Step 20: **Then** air quality is moderate

Step 21: **End if**

Step 22: **Else**

Step 23: Air quality is very severe

Step 24: **End if**

Step 25: **End for**

Step 26: **End**

As given in the above algorithm, to improve the false positive rate or analyze the air quality as it is, not only is high correlation required but also concordance with each other.

Therefore, first, the Least Square Support Vector for the respective selected features is made. Second, to satisfy true positivity, minimum correlation with respect to error should be maintained. Hence, the correlative concordance function employing Kendall's Rank Correlation Coefficient is formulated. With this, air pollution forecasting is done with high accuracy.

#### 4. Experimental setup

Discretized Regression and Least Square Support Vector (DR-LSSV) based air pollution monitoring and control for IoT networks are used to evaluate the proposed method. Several experiments are conducted with Deep AIR [1] and Variational Auto Encoder (VAE) based on the innovative Integrated Multiple Direct Attention Deep Learning architecture (IMDA) (VAE-IMDA) [2]. We also compared the performance of the proposed DR-LSSV method with [1] and [2] in JAVA language using an air quality dataset to measure the efficiency in terms of air pollution forecasting accuracy, air pollution forecasting time and false positive rate.

#### 5. Discussion

This section provides the results analysis of three distinct parameters, air pollution forecasting accuracy, air pollution forecasting time and false positive rate.

##### 5.1. Results Analysis of Air Pollution Forecasting Accuracy

The first set of experiments is conducted to analyze the performance of the proposed method, DR-LSSV, with respect to air pollution forecasting accuracy in forecasting air pollution. With Air Quality Index being applied as the measure by the Indian government to quantify air pollution, air pollution forecasting demands sophisticated monitoring tools and mechanisms along with advanced models to estimate time-related pollutant data. Hence, accurate air quality forecasting is considered to be pivotal for systematic pollution control. The air pollution forecasting accuracy is measured as given below.

$$APF_{acc} = \sum_{i=1}^n \frac{D_{AF}}{D_i} \quad (14)$$

From the above equation (14), the air pollution forecasting accuracy. ‘ $APF_{acc}$ ’ is evaluated based on the number of air quality data involved in a simulation activity. ‘ $D_i$ ’ and the data being accurately forecasted. ‘ $D_{AF}$ ’. It is measured in terms of percentage (%). Table 2 shows the air pollution forecasting accuracy performance comparison of the three methods, DR-LSSV, Deep-AIR [1], and VAE-IMDA [2], based on AQI measurements.

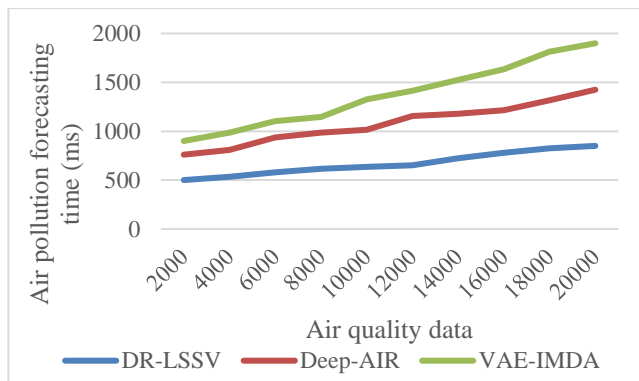


Fig. 5 Performance comparison of air pollution forecasting accuracy

Table 1. Air pollution forecasting accuracy performance comparison of the proposed DR-LSSV method using air quality dataset from India

Air quality data	Air pollution forecasting accuracy (%)		
	DR-LSSV	Deep-AIR	VAE-IMDA
2000	98.25	96.5	95.25
4000	98.05	96	94.25
6000	98	95.75	94.15
8000	97.55	92.15	89.15
10000	97.15	91	87.35
12000	97	90	85.55
14000	96.85	85	82.15
16000	96.25	83.15	78.55
18000	96	82	76.35
20000	95.35	80	72

Figure 5 shows the air pollution forecasting accuracy measured using the three methods DR-LSSV, Deep-AIR [1] and VAE-IMDA [2] considering the AQI measurements. Sample results of ground truth and predicted air pollution forecasting accuracy for the case of air quality data in India are shown in Figure 5. The proposed DR-LSSV method provides air quality data accuracy forecasting via smooth labeling, and it was able to forecast multiple classes based on Kendall’s Rank Correlation Coefficient. The differences between any pair of observations in the output forecasting and ground-truth forecasting for each distinct air quality data ranging between 2000 and 20000 are also provided. On average minimum accuracy differs from 95.35% to 98.25% using DR-LSSV, 80% to 96.5% using [1] and 72% to 95.25% using [2], respectively. With this range, the DR-LSSV method was found to be improved in ensuring accuracy level by 9% compared to [1] and 14% compared to [2]. The reason behind the improvement using the DR-LSSV method was selecting the computationally efficient features using the logistic regression function and maximum semi-parametric likelihood estimator. As a result, with accurate air pollution forecasting, preventive measures in various cities can be taken to control the hazardous factor, resulting in a healthy environment.

##### 5.2. Results Analysis of Air Pollution Forecasting Time

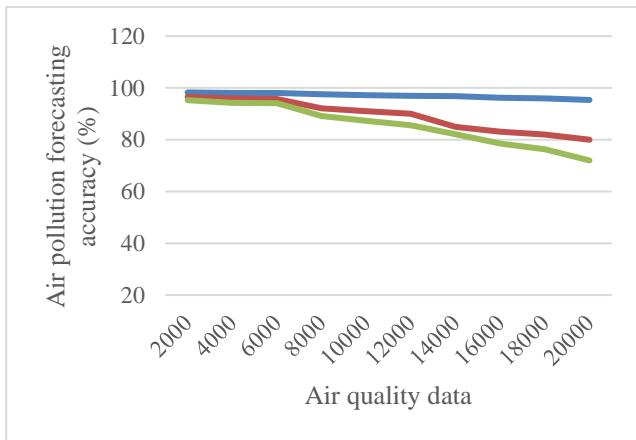
By analyzing historical data sets like air quality data in India, correlations can be made between pollution levels and meteorological data variables, i.e., air quality index. Pollution levels can be measured with the resulting values, ensuring air pollution forecasting efficiently. During this process, a small portion of time is utilized, and that time is referred to as the air pollution forecasting time. The air pollution forecasting time is mathematically formulated as given below.

$$APF_{time} = \sum_{i=1}^n D_i * Time [\tau] \quad (15)$$

From the above equation (15), the air pollution forecasting time. ‘ $APF_{time}$ ’ is measured based on the number of air quality data in concern. ‘ $D_i$ ’ and the time involved in the forecasting process ‘ $Time [\tau]$ ’. It is measured in terms of milliseconds (ms). Table 3 illustrates the air pollution forecasting time comparative results of the considered methods, DR-LSSV, Deep-AIR [1] and VAE-IMDA [2], respectively.

**Table 2. Air pollution forecasting time performance comparison of the proposed DR-LSSV method using air quality dataset from India**

Air quality data	Air pollution forecasting time (ms)		
	DR-LSSV	Deep-AIR	VAE-IMDA
2000	500	760	900
4000	535	810	985
6000	580	935	1105
8000	615	985	1145
10000	635	1015	1325
12000	650	1155	1415
14000	725	1180	1525
16000	780	1215	1635
18000	825	1315	1815
20000	850	1425	1900



**Fig. 6 Performance comparison of air pollution forecasting time**

To study the influence of air pollution forecasting time on distinct numbers of air quality data ranging between 2000 and 20000, experiments were performed with different thresholds (i.e., stations being observed both on a daily and hourly basis). The trends of the forecasting time involved on both a daily and hourly basis can be seen in Figure 6. We note that as the number of air quality data is increased, the percentage of maximum semi-parametric likelihood estimator decreases, and the percentage of air pollution forecasting time decreases steadily. However, the homogenized air quality sample in terms of air quality data increases with the increase in the improperly processed results. This, in turn, results in the minimization of air pollution forecasting time using the DR-LSSV method by 38% compared to [1] and 50% compared to [2], respectively.

**5.3. Results Analysis of the False Positive Rate**

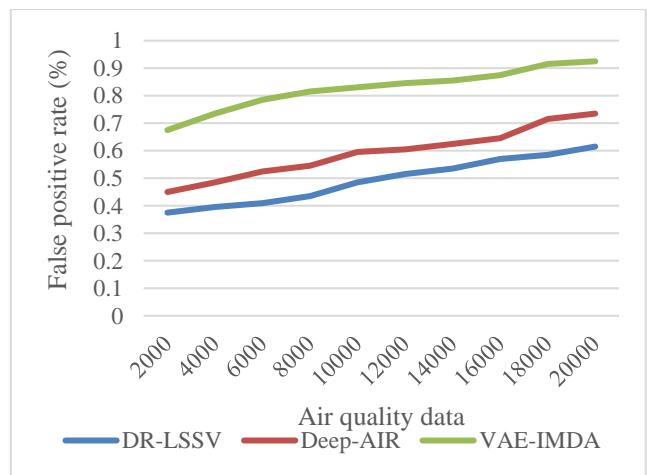
Finally, the third-second set of experiments aims to analyze the performance of DR-LSSV concerning the false positive rate involved in air pollution monitoring. Finally, the false positive rate also referred to as the false alarm ratio, is measured. The false positive rate in air pollution forecasting refers to the probability of falsely rejecting the null hypothesis (i.e., falsely rejecting air pollutants in consideration for measuring the air quality index) for a particular test (i.e., air pollution forecasting). In other words, the false positive rate is measured as the percentage ratio between the number of negative events wrongly categorized as positive (i.e., false air pollutants wrongly involved in the measurement of air quality index) and the total number of air quality data.

$$FPR = \frac{FP}{(FP+TN)} \quad (16)$$

From the above equation (16), the false positive rate ‘ $FPR$ ’ is measured based on the false positive (i.e., wrong inclusion of air pollutants) ‘ $FP$ ’ and the true negatives (i.e., actual wrong air pollutants) ‘ $T$ ’. Finally, table 4, given below, lists the false positive rate.

**Table 3. False positive rate performance comparison of the proposed DR-LSSV method using air quality dataset from India**

Air quality data	False positive rate (%)		
	DR-LSSV	Deep-AIR	VAE-IMDA
2000	0.375	0.45	0.675
4000	0.395	0.485	0.735
6000	0.41	0.525	0.785
8000	0.435	0.545	0.815
10000	0.485	0.595	0.83
12000	0.515	0.605	0.845
14000	0.535	0.625	0.855
16000	0.57	0.645	0.875
18000	0.585	0.715	0.915
20000	0.615	0.735	0.925



**Fig. 7 Performance comparison of false positive rate**

Finally, figure 7 given above illustrates the false positive rate in the y-axis for distinct numbers of air quality data ranging between 2000 and 20000 in the x-axis, respectively. From the above figure, increasing the number of air quality data in the above simulation process results in an increase in the false positive rate using the three methods, DR-LSSV, Deep-AIR [1] and VAE-IMDA [2], respectively. However, with simulations performed using 2000 numbers of air quality data, 15 numbers of air quality data were wrongly assessed using DR-LSSV, 22 numbers air quality data were wrongly assessed using [1], and 27 numbers of air quality data wrongly assessed using [2], the overall false positive rate using the three methods were found to be 0.375, 0.45 and 0.675 respectively. With this result, the false positive rate using the DR-LSSV method was found to be comparatively lesser than [1] and [2]. The reason behind the improvement was the application of the Concordance Correlative Least Square Support Vector-based Classification algorithm in the third hidden layer. Here, multiple classes were evolved according to the resultant values in the AQI based on the processed feature selected. As a result, the falsification of air quality data in the DR-LSSV method was found to be comparatively lesser by 17% compared to [1] and 41% compared to [2].

## 6. Conclusion

Existing methods to forecast air pollution method to detect the level of pollution in the atmosphere using attention mechanism and spatio-temporal features employ 1 to 1 convolution layers from the nonlinear approximations attributes. Our proposed method is capable of obtaining computationally efficient and robust features due to the discretizing of air quality data using sine and cosine functions on both a daily and hourly basis. It automatically extracts strong features from the Constrained Maximum Likelihood function. The Concordance Correlative Least Square Support Vector-based Classification model is applied with the aid of selected features to reduce the false positive rate or falsely select the air pollutants for classification. Finally, multiclass classification results were obtained by means of Kendall's Rank Correlation Coefficient. With this, the DR-LSSV method precisely and accurately forecasts air pollution and minimizes the falsification of air pollutants for further analysis in a timely manner. Experiments on real air quality data in the India dataset show that the DR-LSSV method is superior to the existing method regarding air pollution forecasting accuracy, air pollution forecasting time and false positive rate.

## References

- [1] Qi Zhang et al., "Deep-AIR: A Hybrid CNN-LSTM Framework for Fine-Grained Air Pollution Estimation and Forecast in Metropolitan Cities," *IEEE Access*, vol. 10, pp. 55818-55841, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] Abdelkader Dairi et al., "Integrated Multiple Directed Attention-Based Deep Learning for Improved Air Pollution Forecasting," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-15, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [3] Baowei Wang et al., "Air Quality Forecasting Based on Gated Recurrent Long Short Term Memory Model in Internet of Things," *IEEE Access*, vol. 7, pp. 69524-69534, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] Shengdong Du et al., "Deep Air Quality Forecasting Using Hybrid Deep Learning Framework," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 6, pp. 2412-2424, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] Abdellatif Bekkar et al., "Air-Pollution Prediction in Smart City, Deep Learning Approach," *Journal of Big Data*, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] Shreya Kusrey, Avinash Rai, and Dr. Vineeta (Nigam) Saxena, "Zigbee Based Air Pollution Monitoring and Control System Using WSN," *SSRG International Journal of Electronics and Communication Engineering*, vol. 4, no. 6, pp. 7-11, 2017. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] Yuting Yang, Gang Mei, and Stefano Izzo, "Revealing Influence of Meteorological Conditions on Air Quality Prediction Using Explainable Deep Learning," *IEEE Access*, vol. 10, pp. 50755-50773, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [8] Qing Tao et al., "Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU," *IEEE Access*, vol. 7, pp. 76690-76698, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] Ekta Sharma et al., "Deep Air Quality Forecasts: Suspended Particulate Matter Modeling with Convolutional Neural and Long Short-Term Memory Networks," *IEEE Access*, vol. 8, pp. 209503-209516, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] Mrs.A.Bharathi Lakshmi, Dr.D.Christopher Durairaj, and Mrs.T.Veiluvanthal, "PSNR Based Optimization Applied to Maximum Likelihood Expectation Maximization for Image Reconstruction on a Multi-Core System," *SSRG International Journal of Computer Science and Engineering*, vol. 7, no. 2, pp. 28-41, 2020. [[CrossRef](#)] [[Publisher Link](#)]
- [11] Raquel Espinosa et al., "A Time Series Forecasting Based Multi-Criteria Methodology for Air Quality Prediction," *Applied Soft Computing*, vol. 113, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [12] M. Dhanalakshmi, and V. Radha, "Discretized Linear Regression and Multiclass Support Vector Based Air Pollution Forecasting Technique," *International Journal of Engineering Trends and Technology*, vol. 70, no. 11, pp. 315-323, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

- [13] K. Kumar, and B. P. Pande, "Air Pollution Prediction with Machine Learning: A Case Study of Indian Cities," *International Journal of Environmental Science and Technology*, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [14] Priscilla Whitin, and V. Jayasankar, "A Novel Deep Learning-Based System for Real-Time Temperature Monitoring of Bone Hyperthermia," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 10, no. 1, pp. 187-196, 2023. [[CrossRef](#)] [[Publisher Link](#)]
- [15] Azim Heydari et al., "Air Pollution Forecasting Application Based on Deep Learning Model and Optimization Algorithm," *Clean Technologies and Environmental Policy*, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] Ying Zhang et al., "A Feature Selection and Multi-Model Fusion-Based Approach of Predicting Air Quality," *ISA Transactions*, vol. 100, pp. 210-220, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [17] Yue-Shan Chang et al., "An LSTM-Based Aggregated Model for Air Pollution Forecasting," *Atmospheric Pollution Research*, vol. 11, no. 8, pp. 1451-1463, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] Shakti Chourasiya, and Suvrat Jain, "A Study Review on Supervised Machine Learning Algorithms," *SSRG International Journal of Computer Science and Engineering*, vol. 6, no. 8, pp. 16-20, 2019. [[CrossRef](#)] [[Publisher Link](#)]
- [19] Ichrak Mokhtari et al., "Uncertainty-Aware Deep Learning Architectures for Highly Dynamic Air Quality Prediction," *IEEE Access*, vol. 9, pp. 14765-14778, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [20] Hongqian Chen, Mengxi Guan, and Hui Li, "Air Quality Prediction Based on Integrated Dual LSTM Model," *IEEE Access*, vol. 9, pp. 93285-93297, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [21] Saba Gul, Gul Muhammad Khan, and Sohail Yousaf, "Multi-Step Short-Term *PM* 2.5 Forecasting for Enactment of Proactive Environmental Regulation Strategies," *Environmental Monitoring and Assessment*, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [22] Canyang Guo, Genggeng Liu, and Chi-Hua Chen, "Air Pollution Concentration Forecast Method Based on the Deep Ensemble Neural Network," *Wireless Communications and Mobile Computing*, vol. 2020, p. 13, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [23] Junjuan Xia et al., "Urban Ecological Monitoring and Prediction Based on Deep Learning," *Wireless Communications and Mobile Computing*, vol. 2022, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [24] Qingchun Guo et al., "Air Pollution Forecasting Using Artificial and Wavelet Neural Networks with Meteorological Conditions," *Aerosol and Air Quality Research*, vol. 20, no. 6, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]



**Avinashilingam Institute for Home Science and Higher Education for Women**  
(Deemed to be University Estd. u/s 3 of UGC Act 1956, Category A by MHRD)  
Re-accredited with 'A++' Grade by NAAC.CGPA 3.65/4, Category I by UGC  
Coimbatore – 641 043, Tamil Nadu, India


**PLAGIARISM CHECK REPORT (THESES)**

1.	Name of the Research Scholar	Dhanalakshmi M
2.	Roll No. and Year of Registration	19PHCSF007, 2019
3.	Department	Computer Science
4.	Name of the Research Guide	Dr. V. Radha
5.	Title of the Thesis / Dissertation	Improved Air Pollution Forecasting with Hybrid Machine Learning Algorithms
6.	Similarity Content (%) Identified	9%
7.	Software Used	Turnitin
8.	Date of Verification	06-12-2023


**Note :** The report is excluding 14 Consecutive words, Review of Literature and Quoted Materials.

Checked by :

  
6/12/23  
**Information Scientist**

  
**Research Scholar**

  
06.12.23  
**Assistant Librarian**

  
06/12/23  
**Research Guide**

Date: 06-12-2023

# Improved Air Pollution Forecasting with Hybrid Machine Learning Algorithms

*by Library Adu*

---

**Submission date:** 06-Dec-2023 03:33PM (UTC+0530)

**Submission ID:** 2178355119

**File name:** final\_Thesis\_plagiarism\_file.docx (5.14M)

**Word count:** 33139

**Character count:** 194967

# Improved Air Pollution Forecasting with Hybrid Machine Learning Algorithms

## ORIGINALITY REPORT

9%

SIMILARITY INDEX

6%

INTERNET SOURCES

7%

PUBLICATIONS

1%

STUDENT PAPERS

## PRIMARY SOURCES

1

[arxiv.org](https://arxiv.org)

Internet Source

3%

2

L. V Arun Shalin, A Bharathi, T Prasanth. "Analysis of Multi Cluster Projection on High Dimensional Data Based on Forest Scenario", 2019 International Conference on Advances in Computing and Communication Engineering (ICACCE), 2019

Publication

1%

3

M.Dhanalakshmi, V. Radha. "Air Pollution Forecasting using Novel Feature Selection and Classification Model", international journal of engineering technology and management sciences, 2022

Publication

1%

4

[www.researchgate.net](https://www.researchgate.net)

Internet Source

1%

5

"Data Intelligence and Cognitive Informatics", Springer Science and Business Media LLC, 2023

<1%