PREDICTIVE MONITORING OF HAZARDOUS GASES IN RURAL AND REMOTE AREAS USING MACHINE LEARNING MODELS

E. GANGARAM[1], BODAPATI VIGNESH[2], BUDIDHA HARSHITHA[3], BHUKYA DINESH[4], CHADMALA VINOD[5]

ASSISTANT PROFESSOR[1], UG SCHOLAR[2,3,4&5]

DEPARTMENT OF CSE, NARSIMHA REDDY ENGINEERING COLLEGE (UGC- AUTONOMOUS) MAISAMMAGUDA (V), KOMPALLY, SECUNDERABAD, TELANGANA-500100

ABSTRACT

Monitoring hazardous gases in rural and remote areas is essential for protecting human health, agricultural productivity, and environmental safety. However, traditional gas monitoring systems are often limited by high costs, lack of continuous monitoring, and delayed detection capabilities, especially in geographically isolated regions. To address these challenges, this study proposes a **predictive monitoring system for hazardous gases using machine learning models** that enables real-time detection and early prediction of harmful gas concentrations. The proposed platform integrates **gas sensors, data acquisition units, and machine learning algorithms** to continuously collect and analyze environmental data such as gas concentration levels, temperature, humidity, and atmospheric conditions. The collected sensor data is processed and used to train machine learning models capable of identifying patterns and predicting potential hazardous gas events before they reach critical levels. Algorithms such as Random Forest, Support Vector Machine, or Artificial Neural Networks can be employed to improve prediction accuracy and system reliability. The system also provides **real-time alerts and visualization through a user-friendly monitoring interface**, enabling timely preventive actions by local communities, environmental agencies, and agricultural workers. By combining sensor-based monitoring with intelligent predictive analytics, the proposed approach enhances early warning capabilities and supports safer living conditions in rural and remote environments. Overall, this machine learning–based predictive monitoring platform offers a **cost-effective, scalable, and intelligent solution** for continuous hazardous gas surveillance, contributing to improved environmental monitoring, public safety, and sustainable rural development.

INTRODUCTION

Environmental safety is a growing concern worldwide, especially in rural and remote regions where monitoring infrastructure is limited. Hazardous gases such as carbon monoxide (CO), methane (CH_4), ammonia (NH_3), hydrogen sulfide (H_2S), and sulfur dioxide (SO_2) can pose serious health and environmental risks. These gases may originate from agricultural activities, waste decomposition, industrial emissions, mining operations, or natural processes. In many rural areas, the absence of continuous monitoring systems makes it difficult to detect harmful gas concentrations in time, increasing the risk of health hazards, environmental damage, and accidents. Traditional gas monitoring systems generally rely on periodic manual inspections or standalone sensors that provide limited real-time insights. Such systems often lack predictive capabilities and may fail to provide early warnings before gas concentrations reach dangerous levels. As a result, communities in rural and remote areas remain vulnerable to exposure to toxic gases due to delayed detection and inadequate monitoring infrastructure. Recent advancements in **machine learning (ML)** and **Internet of Things (IoT)** technologies provide an effective solution for intelligent environmental monitoring. Machine learning models can analyze large volumes of sensor data to identify patterns, predict future gas concentration levels, and detect abnormal environmental conditions. By integrating gas sensors with data processing platforms and predictive ML algorithms, it is possible to create a smart monitoring system capable of continuous surveillance and early hazard detection. The proposed system focuses on developing a **predictive monitoring platform for hazardous gases in rural and remote areas using machine learning models**. The system collects environmental data through gas sensors and processes it using ML algorithms to forecast potential gas leaks or harmful concentration levels. This approach enables timely alerts and proactive safety measures, helping protect human health, livestock, and the surrounding ecosystem. By combining sensor-based monitoring with intelligent data analysis, the system aims to enhance environmental safety, improve early warning mechanisms, and support sustainable development in rural and remote communities. This technology-driven approach can play a significant role in building safer environments and reducing the risks associated with hazardous gas exposure.

LITERATURE REVIEW

The monitoring of hazardous gases such as carbon monoxide (CO), methane (CH_4), nitrogen oxides (NOx), ammonia (NH_3), and volatile organic compounds (VOCs) has become an important research area due to their severe impact on human health and environmental safety. Rural and remote regions are particularly vulnerable because of limited monitoring infrastructure, delayed response systems, and lack of continuous environmental surveillance. Recent advancements in **machine learning (ML), Internet of Things (IoT), and sensor technologies** have enabled the development of intelligent systems capable of detecting, predicting, and preventing hazardous gas exposure.

IoT-Based Environmental Monitoring Systems

Early research focused on **IoT-based environmental monitoring systems** that utilize low-cost sensors to measure gas concentration levels in real time. These systems typically use gas sensors such as **MQ-2, MQ-135, and electrochemical sensors** connected to microcontrollers like Arduino or ESP32 to collect environmental data. The collected data are transmitted to cloud platforms for storage and analysis.

Studies on IoT-enabled monitoring demonstrate that integrating sensors with cloud platforms enables continuous tracking of environmental pollutants and supports early warning systems. Advanced environmental monitoring systems have evolved from simple sensing devices to intelligent networks that incorporate IoT communication protocols and cloud computing for data processing and visualization. These systems are widely used for air quality monitoring, pollution analysis, and environmental safety applications.

However, traditional IoT-based monitoring systems mainly rely on **threshold-based detection**, which cannot effectively predict future hazardous gas concentrations or identify complex environmental patterns.

Machine Learning for Gas Detection and Prediction

Machine learning techniques have been introduced to enhance the accuracy and predictive capability of gas monitoring systems. ML models can analyze large volumes of sensor data and identify patterns that indicate potential gas leakage or pollution events.

Research shows that algorithms such as **Random Forest, Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Deep Learning models** can significantly improve the prediction of gas concentrations and environmental trends. These models learn from historical sensor data to forecast hazardous gas levels and trigger alerts before they exceed safety thresholds.

Time-series forecasting models such as **ARIMA, LSTM, and Bi-LSTM** have been widely used for predictive environmental monitoring. Deep learning models, particularly Long Short-Term Memory networks, are capable of capturing temporal dependencies in sensor data, making them suitable for predicting changes in gas concentration levels over time.

Deep Learning and Multimodal Gas Detection

Recent studies have explored the use of **deep learning and multimodal sensing techniques** to improve gas detection accuracy. These approaches combine data from multiple sources such as gas sensors, thermal cameras, and environmental sensors.

Convolutional Neural Networks (CNN) and hybrid models such as **CNN-LSTM and Bi-LSTM architectures** have been applied for gas leakage detection and classification. These models are capable of identifying complex patterns in sensor signals and environmental images, improving the reliability of detection systems. Research has also investigated the use of **federated learning** to enable decentralized training of models while preserving data privacy.

Multimodal gas detection systems provide higher accuracy compared to single-sensor approaches because they integrate different types of environmental data.

Sensor Technology and Machine Learning Integration

Recent advancements in sensor technology have enabled the development of **portable and low-cost gas monitoring systems**. When combined with machine learning algorithms, these sensors can achieve improved sensitivity and specificity in detecting hazardous gases.

Modern sensor arrays, also known as **electronic noses (e-noses)**, use multiple sensors to detect different gas compounds simultaneously. Machine learning algorithms are then used to classify and identify the detected gases based on their unique patterns. This integration improves the reliability of gas detection systems and allows them to operate in dynamic environmental conditions.

Challenges in Hazardous Gas Monitoring Systems

Despite the progress in ML-based monitoring systems, several challenges still exist:

- **Sensor drift and calibration issues** affecting measurement accuracy.
- **Environmental interference** such as temperature, humidity, and dust affecting sensor readings.
- **Limited availability of labeled datasets** for training machine learning models.
- **Connectivity limitations** in rural and remote areas that affect real-time data transmission.
- **High computational requirements** for deep learning models.

These challenges highlight the need for improved data preprocessing, adaptive calibration methods, and hybrid ML models.

Research Gap and Future Directions

Existing research has demonstrated the potential of combining **IoT, gas sensors, and machine learning** for environmental monitoring. However, many current systems still face limitations such as limited prediction accuracy, insufficient scalability, and lack of reliable deployment in rural environments.

Future research directions include:

- Development of **edge-AI systems** for real-time gas prediction.
- Integration of **5G/LoRa communication technologies** for remote data transmission.
- Use of **hybrid ML and deep learning models** for improved prediction accuracy.
- Implementation of **self-calibrating sensor networks** for long-term monitoring.

SYSTEM ANALYSIS

EXISTING SYSTEM

- Hawa'ak used a combination of wireless sensor networks and cellular networks to monitor gases such as O3, CO, and H2S in Qatar, with real-time data visualization and anomaly alerts via Short Message Service (SMS) [31]. In Italy, a pilot project with 300 IoT sensor nodes monitored environmental factors such as C6H6, transmitted data every 30 minutes using solar-powered devices [32]. More recently, the Portuguese prototype (RnProbe) focused on radon gas monitoring [33]. It collected real-time data from IoT devices in rural and urban areas using a combination of Message Queuing Telemetry Transport (MQTT), HyperText Transfer Protocol (HTTP), and Long Range Wide Area Network (LoRaWAN) protocols. Data were visualized via web and mobile platforms, with maps and graphs displaying the radon gas levels. This system also included user authentication and notifications for high radon concentrations, allowing for continuous investigation of radon behavior andits relationship to external causes. Despite their advanced functionality, none of these systems had decision-support tools.

- In 2019, Dhingra et al. developed an external monitoring system called IoTMobair, which has the capability to collect data on pollutant gases such as CO, CO2, NOx , SOx , PM2.5, PM10) [28]. The system uses a Wireless Sensor Networks (WSN) to send data with the HTTP protocol, and the database is hosted in the cloud. The goal of this Indian system is to analyze real-time concentrations of pollutant gases in urban areas, helping people choose less polluted routes during their city commutes. Artificial Intelligence (AI) was used to indirectly measure the Air Quality Index (AQI) based on the environmental variables measured during the pilot phase. AI is commonly used in IoT projects to predict future outcomes and detect patterns in time series data, even from data collected several years ago that was subsequently entered manually [30]. This was demonstrated in the Airtify system [19] and in the system developed by Vanus et al., where AI played a crucial role [20]. In the former case, users were alerted based on the estimated critical zones, and CO2 levels were indirectly predicted.
- In 2021, Vanus et al. presented an improved model that utilized an Artificial Neural Network (ANN) in conjunction with the Bayesian regulation method to recognize patterns [18]. This model was better able to analyze the CO2 levels, temperature, and humidity. The model achieved better results by adding an adaptiveLMSfilter (Sign, Sign-Sign, Sign-Regressor), which improved the accuracy of predicting the CO2 values.
- Ma et al. developed an IoT system similar to this one, which collected data over six months in Northern China to infer real-time data collection [21]. The system monitored pollutant gases, including NO2, SO2, O3, CO, PM1.0, PM2.5, PM10, and the location of the collection device.
- In addition, meteorological data such as temperature, humidity, and wind speed, which can influence the dispersion of gases, were also collected. The data are sent from various devices to a cloud server via cellular networks, which aggregate and store the collected data for visualization. A deep machine learning model, using the Convolutional LSTM (ConvLSTM) algorithm, was applied to handle temporal data for near real- time applications. The developed system visualizes the data using pollution maps, even under irregular sampling conditions, incomplete data, and some climatic variations. The web platform allows any device with Internet access to access real-time and historical data, as well as display environmental pollution maps that are valuable for aiding decision-making and protecting public health. Another ML-based system developed by Rakib et al. integrates IoT and enables the prediction of pollutant gas levels in indoor spaces [23].
- This approach leverages IoT sensors (Temperature, Humidity, Particulate Matter (PMx), Carbon Monoxide (CO), and Ammonia (NH3)) to collect real-time data on environmental conditions and pollutant gases. Subsequently, a pre-trained model was used to predict pollutant gas levels. The ARIMA statistical model was employed to predict time-series data, specifically pollutant gases andAQI for the near future.

DISADVANTAGES

- The complexity of data: Most of the existing machine learning models must be able to accurately interpret large and complex datasets for Monitoring and Prediction of Hazardous Gases.
- Data availability: Most machine learning models require large amounts of data to create accurate predictions. If data is unavailable in sufficient quantities, then model accuracy may suffer.
- Incorrect labeling: The existing machine learning models are only as accurate as the data trained using the input dataset. If the data has been incorrectly labeled, the model cannot make accurate predictions.

PROPOSED SYSTEM

- Development of a platform AI-driven, based on a decision support system for environmental variables and pollutant gases forecasting; Evaluation of ML models based on different metrics of evaluation, taking into account the time series characteristics and trends/seasonality of the data; Implementation and validation of a novel intelligent digital platform for managing air quality in real-time for monitoring houses in rural and remote regions.

ADVANTAGES

In the proposed system, presents the model evaluation and implementation and describes in detail the digital platform for the RuraLTHINGS project. Furthermore, it presents the model validation through areal testbed environment that proved the feasibility of the ML-based platform by successfully predicting data trends related to the indoor monitoring of hazardous gases.

The proposed digital platform and ML model is implemented and part of a IoT project, named RuraLTHINGS. The main goal of this project is to continuously monitor environmental conditions in real time to enhance safety and raise awareness in remote and rural areas. For this proposal, with a robust IoT system, citizens can access all information on multiplatform, which delivers vital information on air quality and issues preventive alerts to improve citizens' quality of life.

IMPLEMENTATION

MODULES

SERVICE PROVIDER

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse Data Sets and Train & Test, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Predicted Hazardous Gases Type Details, Find Hazardous Predicted Gases Type Ratio, Download Predicted Data Sets, View Predicted Hazardous Gases Type Ratio Results, View All Remote Users.

VIEW AND AUTHORIZE USERS

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

REMOTE USER

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, Predict Hazardous Gases Status, VIEW YOUR PROFILE.

MODELS

The proposed digital platform and ML model is implemented and part of a IoT project, named Rural THINGS. The main goal of this project is to continuously monitor environmental conditions in real time to enhance safety and raise awareness in remote and rural areas. For this proposal, with a robust IoT system, citizens can access all information on multiplatform, which delivers vital information on air quality and issues preventive alerts to improve citizens' quality of life.

ALGORITHMS

DECISION TREE CLASSIFIERS

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class

Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

GRADIENT BOOSTING Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.[1][2] When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest.A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

K-NEAREST NEIGHBORS (KNN)

- Simple, but a very powerful classification algorithm
- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not "learn" until the test example is given
- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

Example

- Training dataset consists of k-closest examples in feature space
- Feature space means, space with categorization variables (non-metric variables)
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

LOGISTIC REGRESSION CLASSIFIERS

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar. Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does. This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

NAÏVE BAYES

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature .Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias).While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest of such a technique. Thus, we introduce in a new presentation of the results of the learning process. The classifier is easier to understand, and its deployment is also made easier. In the first part of this tutorial, we present some theoretical aspects of the naive bayes classifier. Then, we implement the approach on a dataset with Tanagra. We compare the obtained results (the parameters of the model) to those obtained with other linear approaches such as the logistic regression, the linear discriminant analysis and the linear SVM. We note that the results are highly consistent. This largely explains the good performance of the method in comparison to others. In the second part, we use various tools on the same dataset (Weka 3.6.0, R 2.9.2, Knime 2.1.1, Orange 2.0b and RapidMiner 4.6.0). We try above all to understand the obtained results.

RANDOM FOREST

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance. The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg. An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision trees with controlled variance. Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

SVM

In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point x and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space. SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to *genetic algorithms* (*GAs*) or *perceptrons*, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

CONCLUSION

Applying machine learning models to IoT devices requires careful adaptation due to variations in sensor data across different environments. The implementation of this precise decision support system has great potential when applied to IoT platforms. The main conclusions and future work of this challenge are found in sections V-A and V-B.

REFERENCES

[1] W. Hu, Y. Wen, K. Guan, G. Jin, and K. J. Tseng, "ITCM: Toward learning-based thermal comfort modeling via pervasive sensing for smart buildings," IEEE Internet Things J., vol. 5, no. 5, pp. 4164–4177, Oct. 2018.

[2] H. Li, S. Li, J. Yu, Y. Han, and A. Dong, "AIoT platform design based on front and rear end separation architecture for smart agricultural,"
in Proc. 4th Asia Pacific Inf. Technol. Conf., 2022, pp. 208–214, doi: 10.1145/3512353.3512384.

[3] B. M. C. Silva, J. J. P. C. Rodrigues, A Ramos, K. Saleem, I. De La Torre, and R. L. Rabělo, "A mobile health system to empower healthcare services in remote regions," in Proc. IEEE

Int. Conf. E-Health Netw., Appl. Services (HealthCom), 2019, pp. 1–6.

[4] World Health Organization (WHO). (2018). 9 Out of 10 People Worldwide Breathe Polluted Air, But More Countries Are Taking Action. Accessed: Nov. 12, 2023. [Online]. Available: https://www.who.int/news-room/detail/02-05-2018-9-out-of-10-peopleworldwide- breathe-polluted-air-but-more-countries-are-taking-action

[5] WHO Handbook on Indoor Radon: A Public Health Perspective, World Health Organization, Geneva, Switzerland, 2009.

[6] N. Rubio-López, A. Llopis-González, Y. Picó, and M. Morales-Suárez-Varela, "Dietary calcium intake and adherence to the Mediterranean diet in Spanish children: The ANIVA study," Int. J. Environ. Res. Public Health, vol. 14, no. 6, p. 637, Jun. 2017, doi: 10.3390/ijerph14060637. [Online]. Available: https://www.mdpi. com/1660-4601/14/6/637

[7] K. Cho, B. van Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in Proc. Conf. Empirical Methods Natural Lang. Process., 2014, pp. 1–15.

[8] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values," Transp. Res. Part C: Emerg. Technol., vol. 118, Sep. 2020, Art. no. 102674, doi: 10.1016/j.trc.2020.102674. [Online]. Available: https://www. sciencedirect.com/science/article/pii/S0968090X20305891

[9] N. Ding, C. Benoit, G. Foggia, Y. Bésanger, and F. Wurtz, "Neural network-based model design for short-term load forecast in distribution systems," IEEE Trans. Power Syst., vol. 31, no. 1, pp. 72–81, Jan. 2016, doi: 10.1109/TPWRS.2015.2390132.

[10] B. T. Ong, K. Sugiura, and K. Zettsu, "Dynamic pre-training of deep recurrent neural networks for predicting environmental monitoring data," in Proc. IEEE Int. Conf. Big Data, Oct. 2014, pp. 760–765, doi: 10.1109/BIGDATA.2014.7004302.

[11] R. P. Masini, M. C. Medeiros, and E. F. Mendes, "Machine learning advances for time series forecasting," J. Econ. Surv., vol. 37, no. 1,
pp. 76–111, Feb. 2023, doi: 10.1111/joes.12429.

[12] J. Crespo Cuaresma, J. Hlouskova, S. Kossmeier, and M. Obersteiner, "Forecasting electricity spot-prices using linear univariate time-series models," Appl. Energy, vol. 77, no. 1, pp. 87–106, Jan. 2004, doi: 10.1016/s0306-2619(03)00096-5. [Online]. Available: https://www. sciencedirect.com/science/article/pii/S0306261903000965

[13] G. P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," Neurocomputing, vol. 50, pp. 159–175,
Jan. 2003, doi: 10.1016/s0925-2312(01)00702-0. [Online]. Available:
https://www.sciencedirect.com/science/article/pii/S0925231201007020

[14] R. Singhal, N. K. Choudhary, and N. Singh, "Short-term load forecasting using hybrid ARIMA and artificial neural network model," in Advances in VLSI, Communication, and Signal Processing. Singapore: Springer, D. Dutta, H. Kar, C. Kumar, and V. Bhadauria, 2020, pp. 935–947, doi: 10.1007/978-981-32-9775-3_83.

[15] Q. Tao, F. Liu, Y. Li, and D. Sidorov, "Air pollution forecasting using a deep learning model based on 1D ConvNets and bidirectional
GRU," IEEE Access, vol. 7, pp. 76690–76698, 2019, doi: 10.1109/ACCESS.2019.2921578.