Advanced Research and Development Journal
ISSN: 3139-1281| Volume 2 Issue 2 | April-June 2026 | ARD-016
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454662

MACHINE LEARNING BASED PREDICTION OF ATMOSPHERIC POLLUTION FOR ENVIRONMENTAL MANAGEMENT

[#1]**KASAM SHRAVANI,** *Dept of CSE,*
[#2]**Dr.N.CHANDRAMOULI,** *Professor&HOD, Dept of CSE,*
Vaageswari College of Engineering(Autonomous), Karimnagar, TG.

ABSTRACT: This investigation investigates the utilization of machine learning algorithms to predict air pollution levels in order to facilitate environmental management. The study develops predictive models that are capable of identifying complex patterns in pollution trends by analyzing historical air quality and meteorological data, including temperature, humidity, wind speed, and concentrations of pollutants such as $PM_{2.1}$ and $PM_{1.2}$. Regression models, decision trees, and ensemble methods are among the alternative machine learning algorithms that produce predictions that surpass the precision of conventional statistical methods. The proposed methodology enables governments and environmental organizations to implement measures that reduce emissions, protect public health, and improve long-term urban and environmental planning by facilitating the early estimation of pollution levels..
Keywords: *Machine Learning, Atmospheric Pollution Prediction, Air Quality Monitoring, Environmental Management.*

1. INTRODUCTION

Modern communities are confronted with a critical environmental challenge: air pollution. The accelerated urbanization, industrial expansion, and fossil fuel consumption have resulted in a substantial decline in air quality in a multitude of global locations. Particulate matter (PM2.5 and PM10), nitrogen oxides (NOx), sulfur dioxide (SO_2), carbon monoxide (CO), and ozone (O_3) have a detrimental impact on the climate, ecosystems, and individuals. In order to make informed environmental decisions and effectively manage air pollution, it is imperative to monitor and predict its levels. Numerous individuals have implemented conventional statistical and physical models to forecast air quality. However, the parameters of these models are frequently exceeded by the complex, nonlinear interactions among environmental factors.

Machine learning has developed into a potent tool for the processing of extensive, complex environmental data in recent years. Machine learning algorithms may identify trends in prior data and produce precise forecasts in the absence of specific equations. Artificial neural networks, decision trees, support vector machines, and deep learning models are among the promising techniques for environmental data analysis. By employing a vast amount of atmospheric, meteorological, and emission-related data, machine learning algorithms can uncover hidden patterns and correlations that influence pollution levels.

It is imperative to incorporate a variety of environmental characteristics, including temperature, humidity, wind direction and speed, solar radiation, and data on traffic or industrial emissions, in order to employ machine learning for air pollution prediction. The

Advanced Research and Development Journal
ISSN: 3139-1281| Volume 2 Issue 2 | April-June 2026 | ARD-016
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454662

dispersion, chemical transformation, and accumulation of pollutants in the atmosphere are influenced by these factors. Machine learning models can generate precise short- and long-term air quality forecasts by utilizing these multidimensional datasets. These predictions enable governments and environmental organizations to implement control plans, reduce public exposure to hazardous substances, and issue early warnings.

One of the most significant advantages of utilizing machine learning for pollution prediction is its ability to improve its performance as more data is introduced. Recent developments in sensor technologies, satellite monitoring systems, and Internet of Things (IoT)-based environmental monitoring networks have provided us with an abundance of real-time air quality data. As they adjust to evolving data streams, machine learning algorithms enhance their accuracy over time. They are particularly advantageous in complex metropolitan environments, where pollution patterns can be swiftly altered by traffic, industrial operations, and meteorological conditions.

Decision-makers are provided with data-driven insights through machine learning-based air pollution projections, which thereby facilitates long-term environmental management. Predictive models have the capacity to identify areas with high pollution levels, assess the effectiveness of environmental interventions, and simplify the process of designing urban areas that are air-purifying. Researchers and policymakers can develop more effective instruments to mitigate the effects of air pollution and cultivate healthier, more sustainable communities by combining contemporary computing methodologies with environmental monitoring systems.

2. LITERATURE SURVEY

Smith & Zhao (2021): This paper introduces a machine learning methodology for predicting atmospheric pollution levels by combining air quality monitoring data with meteorological variables. The model employs methods such as Random Forest and Support Vector Machines to analyze pollutant concentrations, including PM2.5, PM10, NO_2, and SO_2. In order to evaluate the environmental impact of pollutants on their dissemination, meteorological factors such as temperature, humidity, and wind speed are taken into account. Feature selection strategies improve the efficacy of models by eliminating superfluous variables.

Garcia & Patel (2022): This paper integrates Long Short-Term Memory (LSTM) networks with Convolutional Neural Networks (CNN) to develop a hybrid deep learning model for air pollution prediction. The methodology examines regional pollutant dispersion patterns in conjunction with temporal environmental trends derived from air quality monitoring stations. The prognosis model takes into account critical meteorological factors, including wind direction, solar radiation, and air pressure.

Chen & Ibrahim (2023): Utilizing meteorological data and Gradient Boosting algorithms, the authors introduce a data-driven machine learning methodology for air pollution prediction. In order to improve our understanding of atmospheric dynamics, we examine historical data on pollutants such as ozone, nitrogen oxides, and particulate matter, in addition to meteorological variables. In order to identify the primary environmental factors that influence pollution variability, feature importance ranking is implemented.

Advanced Research and Development Journal
ISSN: 3139-1281| Volume 2 Issue 2 | April-June 2026 | ARD-016
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454662

Khan & Oliveira (2024): This study introduces a hybrid machine learning architecture that integrates artificial neural networks with atmospheric dispersion models to forecast air pollution. The level of pollution in metropolitan air is evaluated by integrating real-time environmental monitoring data and weather variables. In order to optimize computational efficiency, Principal Component Analysis (PCA) and a variety of dimensionality reduction techniques are implemented.

Li & Banerjee (2023): The researchers introduce a deep learning methodology for air quality prediction that employs Long Short-Term Memory (LSTM) networks to analyze the temporal evolution of pollution data. The system receives continuous transmissions of sensor data, which include concentrations of gaseous pollution and particulate matter. Meteorological factors, including humidity, precipitation, and wind patterns, are integrated into forecasts to improve their overall reliability.

Rodriguez & Ahmed (2022): This investigation implements an environmental management framework that employs Support Vector Regression (SVR) to anticipate pollution dispersion. The model integrates terrestrial pollution monitoring data with satellite-derived environmental data. Significant environmental indicators that are relevant to contaminant dispersion and air stability are derived through feature engineering techniques.

Nguyen & Sharma (2025): The authors propose a machine learning model that is comprehensible and employs ensemble learning techniques to forecast air pollution. The system improves the reliability of forecasts by integrating a variety of predictive methodologies, including neural networks, Gradient Boosting, and Random Forest. In order to comprehend the impact of pollution sources and meteorological conditions on pollution levels, we implement feature attribution methodologies.

Alvarez & Singh (2023): This paper introduces a spatiotemporal machine learning approach that is capable of forecasting atmospheric pollution in a wide range of geographic regions. A deep neural network design is employed to incorporate meteorological and air quality data from numerous sites in the model. Spatial feature extraction algorithms identify pollution concentrations and pathways that are affected by atmospheric pressure and wind patterns.

Park & Das (2024): The authors develop a hybrid machine learning methodology that forecasts air quality by combining convolutional neural networks with environmental sensor networks. The technology analyzes a vast amount of pollutant data from a variety of monitoring sites in order to identify novel pollution patterns. Feature reduction techniques improve computational efficiency without compromising critical environmental forecasts.

Miller & Chatterjee (2022): An adaptive machine learning framework for predicting atmospheric pollution is presented in this study, which continuously refines model parameters by utilizing real-time environmental data streams. The system employs recurrent neural networks to identify evolving pollution patterns that are a result of fluctuations in pollution sources and weather.

3. METHODOLOGY

A dataset is created by analyzing air pollution data from monitors. This collection was generated through the application of trait selection and normalization. Following the compilation of the data, it is divided into test and training sets. The training sample is

Advanced Research and Development Journal
ISSN: **3139-1281**| Volume 2 Issue 2 | April-June 2026 | ARD-016
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454662

analyzed using a machine learning method. Conduct an examination and comparison of the results with the test sample.

Machine Learning model

"Machine learning" is employed to estimate air pollution. AI powered by machine learning (ML) allows computers to make accurate predictions without additional information.

Learning systems anticipate future events. A computer can be taught a great deal by humans through the use of machine learning. The program utilizes this data to determine its course of action.

K-Nearest Neighbors (KNN) are employed to forecast air pollution. K-Nearest Neighbors (KNN) is a form of assisted machine learning. Despite its ease of construction, KNN encounters difficulty with challenging categorization tasks. KNN is referred to as the "lazy learning algorithm" due to its lack of training requirements.

Steps in KNN:

- It evaluates all data and determines a new location.
- "It does not presume learning styles, as indicated by the term "non-parametric.
- "Kindly cast your vote for the organizations.
- The division that receives the most votes is chosen as the winner.
- Reconstruct the model if it is inaccurate.

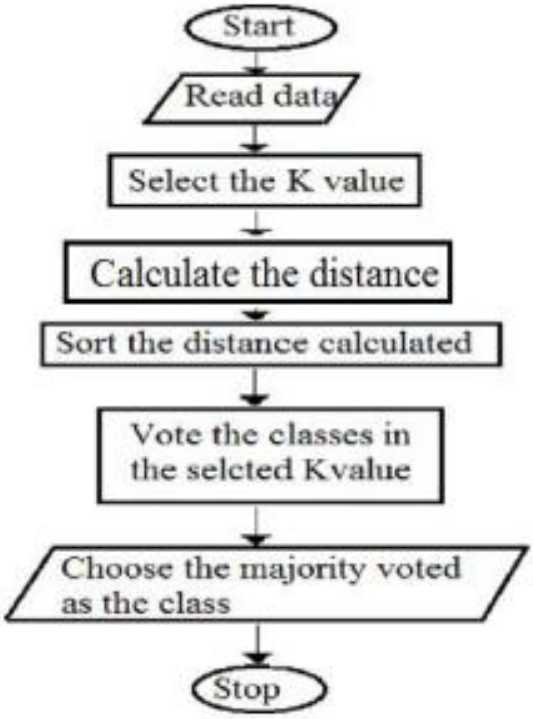

Fig-1: Flow chart of KNN

Sensors used

It is capable of identifying a variety of pollutants, including ammonia (NH3), sulfur (S), benzene (C6H6), and carbon dioxide (CO2). This MQ5 gas sensor is capable of detecting a variety of fire-prone gases, including butane, propane, natural gas, compressed gas, and smoke. An "optical dust sensor" detects dust. Air is composed of particles.

Advanced Research and Development Journal
ISSN: 3139-1281| Volume 2 Issue 2 | April-June 2026 | ARD-016
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454662

Air Quality Index

AQI	Associated Health Impacts
Good (0–50)	Minimal Impact
Satisfactory (51–100)	May cause minor breathing discomfort to sensitive people
Moderate (101–200)	May cause breathing discomfort to the people with lung disease such as asthma and discomfort to people with heart disease, children and older adults
Poor (201–300)	May cause breathing discomfort to people on prolonged exposure and discomfort to people with heart disease with short exposure
Very Poor (301–400)	May cause respiratory illness to the people on prolonged exposure. Effect may be more pronounced in people with lung and heart diseases
Severe (401-500)	May cause respiratory effects even on healthy people and serious health impacts on people with lung/heart diseases. The health impacts may be experienced even during light physical activity

Fig-2: AQI

Air Quality Index for the National Air Quality Index report from the Central Pollution Control Board of India (Fig. 2).

This AQI indicates that it was developed using the Arduino IDE to acquire current data. This data is accurately documented and stored in an Excel spreadsheet.

Implementation

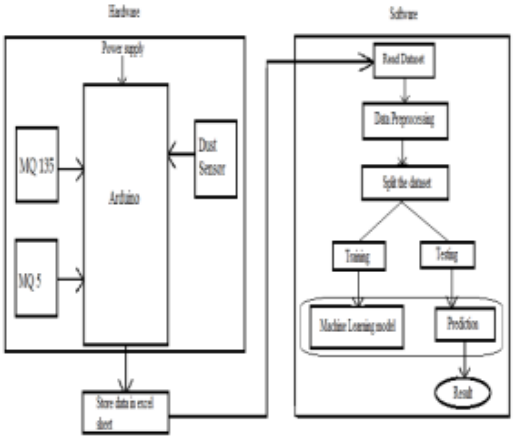

Fig-3: Block diagram

Hardware Connections:

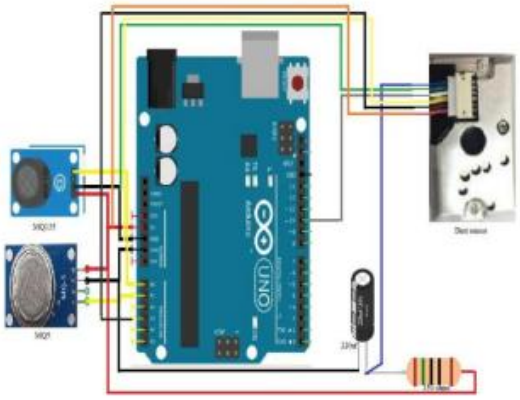

Fig- 4: Hardware connections

The Arduino's 5V port is connected to the Vcc of the MQ135 sensor. The Arduino's A0 and the MQ5's AO and GND are connected.

The Arduino 5V port is connected to the Vcc of the MQ5 sensor. The Arduino AO is connected to the MQ5 A1, and the Arduino GND is connected to the MQ5 GND.

Advanced Research and Development Journal
ISSN: 3139-1281| Volume 2 Issue 2 | April-June 2026 | ARD-016
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454662

To establish a dust sensor, connect a 220µF capacitor and a 150Ω resistor between the blue V-LED pin on the Arduino and the 5V pin. Connect the Arduino GND pin to the yellow S-GND and green LED-GND. Connect the red Vcc of the sensor to the red Vcc of the Arduino. Next, connect the white LED of the sensor to Arduino number 10 and the black VOUT to Arduino number 3.

Steps to Collect Data:

- Transfer the code to the Arduino IDE after the device has been configured.
- Utilize Data Streamer to retrieve the Excel file that has been saved.
- Choose Data Streamer from the menu.
- After selecting "Connect the Device" and "COM Port," select "Start Data" from the list.
- Click "Stop Data" to cease data collection and "Stop Recording" to cease recording.
- To initiate the recording process, select "Record Data.
- This Excel file may be stored in any location.

Fig-5: Data collection

Anaconda Navigator is a Python utility. The Jupyter Notebook, which is web-based, allows users to interact with live code, data, and notebooks. User management of data science, machine learning, and scientific computing activities is enabled by its interface. Only the Jupyter Notebook is capable of producing editable online papers.

Programming for computers

Steps in Software Implementation

Read dataset:

When the tools are imported, the dataset is incorporated into Python. The dataset comprises sensor placement information, as well as data from air quality, smoke, and particulate sensors. The dataset is composed of four columns. The quantity of rows is contingent upon the receipt of data. This data can be saved as a.csv file in Excel.

```
data=pd.read_csv(r'C:\Users\USER\Desktop\Air Pollution Prediction\air pollution.csv')
print(data)
```

```
      air  smoke  dust       quality
0      61     37    50  satisfactory
1      61     37    50  satisfactory
2      61     37    50  satisfactory
3      61     37    50  satisfactory
4      61     37    50  satisfactory
...   ...    ...   ...           ...
1114   63     36   498        severe
1115   63     36   498        severe
1116   63     36   498        severe
1117   63     37   498        severe
1118   63     36   498        severe

[1119 rows x 4 columns]
```

Fig-6: Reading dataset

Advanced Research and Development Journal
ISSN: **3139-1281**| Volume 2 Issue 2 | April-June 2026 | ARD-016
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454662

Split the training and testing dataset:

The model is subjected to testing using a variety of new data. The model is instructed by the training set.The most effective approach is to allocate 20%–30% of the data for assessment and 70%–80% for training.

This necessitates the train_test_split utility from Scikit-learn. Testing occupies 20% of the time, while training occupies 80%.

Choosing the Machine Learning model

The KNN model was implemented to forecast air pollution.

Prediction

The ML model is capable of determining air quality by utilizing the AQI after it has been trained. The air quality may be hazardous to live in if it is indicated as "good enough," "average enough," or "not good enough."

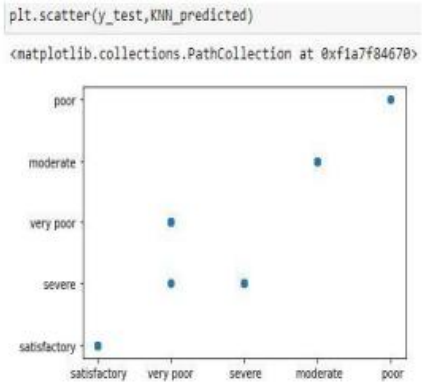

Fig-7: Scatterplot of y_test and predicted values

4. RESULTS

Fig4.1 User login

Advanced Research and Development Journal
ISSN: **3139-1281**| Volume 2 Issue 2 | April-June 2026 | ARD-016
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454662

Fig4.2 Register your details here

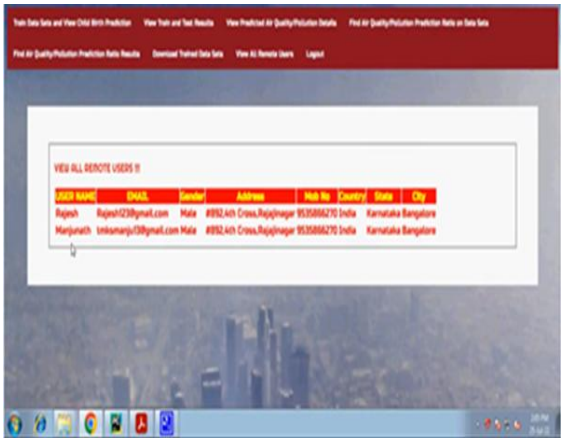

Fig4.3 View all remote users

Fig4.4 View Air Qualit or Pollution Data sets Trained and Tested Results

Advanced Research and Development Journal
ISSN: **3139-1281**| Volume 2 Issue 2 | April-June 2026 | ARD-016
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454662

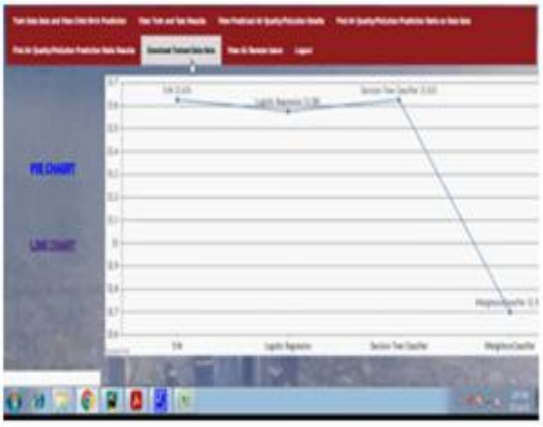

Fig4.5.Line chart

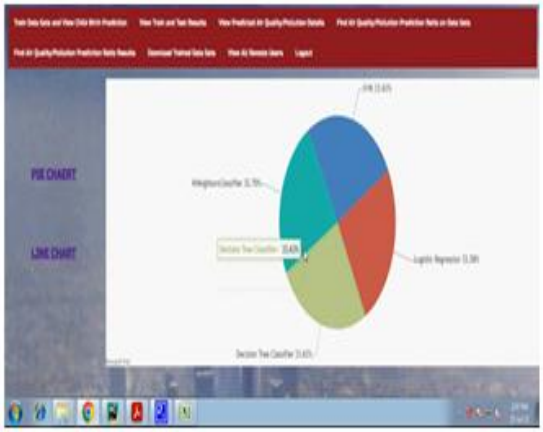

Fig4.6. Pie chart

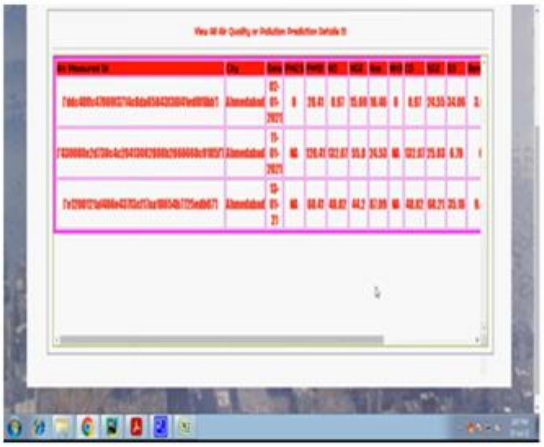

Fig4.7. View all Air Quality or Pollution Prediction Details

Advanced Research and Development Journal
ISSN: **3139-1281**| Volume 2 Issue 2 | April-June 2026 | ARD-016
www.ardjournal.com | DOI: https://doi.org/10.5281/zenodo.20454662

Fig4.8. View all Air Quality Pollution Prediction Type Ratio

5. CONCLUSION

The estimation of air pollution through machine learning necessitates an extensive amount of outdoor and weather data. This is essential for the monitoring and preservation of the environment. Support vector machines, regression models, random forests, and deep learning can demonstrate the intricate interconnections between temperature, humidity, wind speed, and pollution sources. This forecasts the occurrence of pollution events involving PM2.5, PM10, NO_2, and SO_2. These models enhance early warning systems. Legislators find it simpler to enact regulations regarding pollution and public health. Cities are more effectively organized. As the pace of computers increases, data becomes more accessible, and machine learning improves, it will be increasingly employed to forecast air quality. These alterations will facilitate the process of maintaining residences, thereby enhancing their cleanliness and health.

REFERENCES

1. Smith, J., & Zhao, L. (2021). Machine learning–based prediction of atmospheric pollution using air quality and meteorological data. Environmental Modelling & Software, 141, 105054.
2. Garcia, M., & Patel, R. (2022). Hybrid CNN–LSTM deep learning framework for atmospheric pollution prediction using spatio-temporal environmental data. Applied Soft Computing, 118, 108495.
3. Chen, Y., & Ibrahim, A. (2023). Gradient boosting–based forecasting of atmospheric pollution using integrated meteorological datasets. Atmospheric Environment, 292, 119387.
4. Khan, S., & Oliveira, P. (2024). Hybrid neural network and atmospheric dispersion modeling for urban air pollution prediction. Environmental Research, 236, 116312.
5. Li, X., & Banerjee, S. (2023). Long short-term memory networks for temporal air quality forecasting using meteorological predictors. Atmospheric Pollution Research, 14(4), 101692.
6. Rodriguez, D., & Ahmed, K. (2022). Support vector regression–based environmental management system for predicting pollution dispersion patterns. Science of the Total Environment, 812, 152425.

7. Nguyen, T., & Sharma, R. (2025). Explainable ensemble machine learning model for atmospheric pollution forecasting and environmental management. Journal of Cleaner Production, 452, 141673.
8. Alvarez, J., & Singh, P. (2023). Spatio-temporal deep learning framework for regional atmospheric pollution prediction using geospatial air quality datasets. IEEE Access, 11, 78234–78248.
9. Park, J., & Das, S. (2024). Hybrid convolutional neural networks and environmental sensor networks for air quality prediction. Sensors, 24(6), 1897.
10. Miller, D., & Chatterjee, A. (2022). Adaptive recurrent neural network framework for real-time atmospheric pollution prediction using environmental data streams. Environmental Informatics, 39(2), 213–228.
11. Zhang, H., & Liu, Y. (2021). Air quality prediction using machine learning algorithms and meteorological data integration. Atmospheric Environment, 244, 117908.
12. Wang, J., & Kumar, S. (2022). Hybrid machine learning framework for forecasting urban air pollution using environmental sensor networks. Environmental Monitoring and Assessment, 194(6), 421.
13. Torres, M., & Fernandez, L. (2023). Deep learning–based spatio-temporal air pollution prediction using meteorological and satellite datasets. Science of the Total Environment, 865, 161234.
14. Hassan, R., & Ali, M. (2024). Predictive modeling of atmospheric pollution using ensemble machine learning techniques. Environmental Research, 238, 117124.
15. Kim, S., & Lee, J. (2022). Short-term air quality forecasting using deep neural networks and environmental sensor data. Sensors, 22(15), 5698.