Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2025 SCPE. Volume 26, Issues 3, pp. 1005–1016, DOI 10.12694/scpe.v26i3.4190

SMART AGRICULTURE: INTEGRATING AIR QUALITY MONITORING WITH DEEP LEARNING FOR PROCESS OPTIMIZATION*

SHOBANA J† VENKATA SUBRAMANIAN A‡ BALAMURUGAN P § SIVAKUMAR PERUMAL ¶ SANKARI V‖ ELDHO K J** AND NARESHKUMAR R††

Abstract. Modernization and intense industrialization have led to a substantial improvement in people's quality of life. However, the aspiration for achieving an improved quality of life results in environmental contamination. A primary consequence of environmental degradation is air pollution, resulting from rising levels of poisonous chemicals in the atmosphere, which may induce detrimental health conditions in humans. It is harmful to both humans and agriculture. Given that the effects of air pollution on plants may not be readily apparent, it is important to analyse the necessary data and compute the outcomes. Farmers prioritise on pests and plant diseases, frequently neglecting the detrimental impacts of air pollution. Some plant species can withstand high amounts of pollution from suspended particulate matter and accumulated gases, while others are more susceptible to harm. Therefore, plants' reaction to air pollution is influenced by the kind of harmful compounds, their levels, and the plant's susceptibility to them. The LSTM +CNN Proposed Ensemble method may be used to analyse the impact of air pollution on agriculture by examining trends in crop production over time and predicting which crop is more resistant based on the pollution data. The initiative created for this aim may assist farmers in determining the most suitable crop to cultivate in their fields to minimize the impact of air pollution on agricultural yield. The findings show deep learning algorithms correctly predict hourly pollutant concentrations such as carbon monoxide, sulphur dioxide, nitrogen dioxide, ground-level ozone, and particulate matter 2.5, along with the hourly Air Quality Index (AQI) for California. A proposed model used test RMSE values as a measure to evaluate prediction performance, achieving the best possible results.

Key words: Air quality, Deep Learning, Pollution Environmental, neural networks.

1. Introduction. The effects of air pollution on agricultural productivity are not readily apparent until closely watched and analysed. Farmers often deal with challenges related to irritations and infections in their plants. Even though the aforementioned problem is being addressed, the detrimental impacts that are brought about by air pollution are not being addressed. The knowledge that is anticipated to differentiate and carry out these advancements is not practical nor available at a practicable level, even though the modifications in agricultural techniques may reduce the severity of these consequences[1].

Plant species that are distinct from one another react differently to pollution. Even though certain plant species may be able to withstand critical levels of pollution brought on by suspended particulate matter and buildup gases, other plant species are often susceptible to harm. As a result, how plants react to air pollution is contingent upon the many kinds of harmful compounds that are present, the quantity of those pollutants, and the extent to which they are susceptible to them [2].

*This work was supported by the Society for Industrial and Applied Mathematics

†Department of Data Science and Business Systems, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Tamilnadu, India (shobanaj1@srmist.edu.in).

‡Department of Computer Science and Engineering, GITAM University, Bengaluru, India (vadhinar@gitam.edu).

§Department of Computer Science and Engineering, J N N Institute of Enginering, India and Saveetha university, Chennai, India (balamuruganp@jnn.edu.in).

¶Department of Computer Science and Engineering, Sasi Institute of Technology and Engineering, Tadepalligudem, Andhra Pradesh, India (sivakumarperumal@sasi.ac.in).

‖Department of Artificial Intelligence and Data Science K.Ramakrishnan College of Engineering, Samayapuram - Kariyamanickam Rd, Tamil Nadu, India. (sankarime09@gmail.com).

**PG and Research Department of Computer Science, Mary Matha Arts and Science College, Mananthavady, Kerala, India. (eldhokj@marymathacollege.ac.in).

††Department of Networking and Communications, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Tamilnadu, India. (nr7061@srmist.edu.in).

Urban areas are experiencing environmental pollution issues such water, noise, and air pollution due to economic and technological advancements. Air pollution directly affects human health by exposing individuals to toxins and particles, leading to a growing interest in its effect among scientists. The primary sources of air pollution are the combustion of fossil fuels, agricultural activities, emissions from factories and businesses, household heating, and natural calamities [2]. Air quality in the United States has been researched for the last thirty years with the establishment of the Clean Air Act programme. Despite the program's improvement in air quality over time, air pollution remains a persistent issue. About 200,000 premature deaths per year in the US are attributed to total combustion emissions, mostly caused by pollutants like particulate matter 2.5 (PM2.5), with an additional 10,000 fatalities per year linked to variations in ozone concentration.

The American Lung Association estimates that air pollution-related diseases cost about 37 dollars billion annually in the US, with California accounting for 15 billion dollars [7].Air pollution is the introduction of dangerous or excessive amounts of certain chemicals like gases, particles, and biological molecules into the atmosphere.

The high emissions result in detrimental effects such as infections, fatalities among people and other living species, and damage to crops. The primary air pollutants, referred to as criteria pollutants, include CO, SO2, lead, ground-level ozone (O3), NO2, and PM. The US Environmental Protection Agency (EPA) monitors the levels of these pollutants to regulate air quality. Scientific studies have shown a connection between brief exposure to these pollutants and various health issues, such as reduced capacity to meet higher oxygen requirements during physical activity (especially for individuals with heart conditions), airway inflammation in healthy individuals, heightened respiratory symptoms in asthma patients, respiratory crises in children and the elderly, and more.

The influence of air pollution on agricultural output is substantial, although frequently disregarded. While farmers tackle plant irritations and infections, the wider detrimental impacts of air pollution remain unattended to. Various plant species exhibit distinct responses to pollution, with certain species being more vulnerable to harm caused by pollutants such as particulate matter and gases. Urban areas are exposed to pollution from multiple sources, such as the burning of fossil fuels, emissions from industries, and agricultural practices. This pollution has a negative impact on human health and leads to premature deaths. Despite the implementation of initiatives such as the Clean Air Act, air pollution continues to be a significant problem. Machine learning (ML) provides potential solutions for assessing and reducing these effects by extracting valuable information from extensive datasets to forecast and enhance agricultural results.

Machine Learning(ML) is a subset of artificial intelligence(AL) that involves using statistical models to extract important insights from large datasets. The main difference between the method of statistical analysis and machine learning is that statistics focuses on representing data in numerical terms of probability or likelihood measures, rather than deterministic processes like cluster assignments, forecasting functions, etc. The assignments and tasks to be completed are roughly equivalent. The learning techniques are referred to as estimating strategies. Many researchers and analysts have already found that the basic concepts of machine learning closely resemble non-parametric estimation terminology [5]. ML allows systems to learn and improve from data without explicit programming. ML is distinct from AL and DL. ML's benefit is in training the model based on available data to predict future outcomes more efficiently. The subsections detail the individual characteristics of several supervised and unsupervised algorithms that impact agricultural frameworks.

The paper is structured as outlined below. Section 2 provides a thorough review of the literature, analysing past and relevant research. In Section 3, we present a suggested deep learning model, emphasising its predictive capabilities. Furthermore, provide a detailed description of the data used in this study and elaborate on the data pretreatment steps taken to create a more concise and informative dataset for analysis. Section 4 outlines our experimental investigation, including details of the experimental setup and analysis of the data. Section 5 closes the work and presents suggestions for further research.

2. Related Work. Saritha et al [3] the article underscores the harmful impact of air pollution on agriculture, noting that while the impacts may not be immediately apparent, they may greatly reduce agricultural productivity. The authors suggest using a machine learning method to examine the impact of air pollution on agriculture, focusing on the trends in crop production over time. This technique seeks to identify crop types that are more resistant by analysing pollution data, enabling farmers to make well-informed choices when selecting crops. The research explores the integration of air quality data with agricultural yield data to analyse the impact of contaminants on crop production. It indicates that certain crops have decreased production when exposed to high levels of pollutants, but others, such as sugarcane, are more resilient. The protocols offer an overview of available supervised and unsupervised machine learning models [4] connected with agricultural yield in literature. Highlighting the promise of machine learning in tackling difficult agricultural concerns including crop improvement, yield prediction, crop disease diagnosis, and recognizing water stress. Exploring the combination of agronomic elements with data analytic approaches to enhance crop yield forecasting. This might be helpful for agricultural academics and practitioners who want to use data-driven methods to enhance crop management and productivity.

The research [5] combines support vector regression (SVR) with a radial basis function (RBF) kernel to successfully estimate the concentrations of pollutants and the air quality index (AQI) in California. The authors show how employing the complete set of accessible variables is more successful than feature selection utilizing principal component analysis. The report offers future research areas, such as examining additional approaches for hyperparameter optimization and comparing SVR findings with additional algorithms for machine learning, which may lead to further breakthroughs in air quality prediction and simulation.

A complete survey of the current achievements in the application of deep learning (DL) in the agriculture industry. Highlights numerous uses of DL in agriculture, covering counting fruits, controlling water, crop management, soil management, weed identification, seed categorization, yield prediction, disease detection, and harvesting [6]. It underlines the problems encountered in applying DL in agriculture, including the complexity of assembling datasets, the expense of processing resources, and the scarcity of DL professionals. Addressing these difficulties may help overcome hurdles to the mainstream implementation of DL in agriculture.

The report provides an in-depth examination of recent developments and identifies areas for further investigation, such as robustness, interpretability, and integration of multiple data modalities. As such, it is an invaluable resource for future research and development in the field of deep learning in agriculture.

The study's methodology included a thorough examination of agricultural deep learning algorithms by analysing secondary data from academic publications released between 2016 and early 2022. Data collecting included using databases including Research Gate, IEEE Explore, Springer, Elsevier, Google Scholar, Frontier, and Science Direct. The focus was on scholarly journal articles and conference papers that were pertinent to the study goals. Studies predating 2016 were not included in the research. The paper utilised a range of deep learning tools for agricultural model development, such as Python tools for image saliency, gradient explanation technique, integrated gradient, DeepLIFT, guided backpropagation, class activation maps (CAMs), and layer-wise relevance propagation (LRP). The technologies were used to improve the precision and comprehensibility of deep learning models in the field of agriculture.

Assessing the current status of agricultural air quality research and pinpointing potential future research avenues to investigate contaminants associated with agriculture and their effects on air quality, human health, and regional climate. Developments in evaluations, modelling, emission controls, and farm operation management are necessary to successfully limit emissions from agriculture [7]. The significance of implementing laws and regulations to decrease agricultural emissions and their environmental effects. It is important to tackle the issues and uncertainties in present air quality models used in agriculture since doing so would enhance air quality, human health, agricultural settings, and biodiversity.

Examining recent studies and uses of artificial intelligence to lessen the negative impacts of climate change, particularly in fields like energy efficiency, carbon capture and storage, weather and renewable energy prediction, grid control, architectural design, transportation, precision farming, industrial operations, deforestation reduction, and sustainable urban development.AI can play a crucial role in mitigating the effects of climate change by improving energy efficiency, decreasing energy usage in buildings, and optimizing power systems to lower electricity costs [8]. Integrating AI with smart grids can enhance the efficiency of power systems, resulting in less energy wastage and reduced electricity costs. AI integrated with transportation systems can decrease carbon dioxide emissions by around 60 percent. AI can assist in the conservation of natural resources by decreasing deforestation and encouraging sustainability. It can also help in designing resilient cities to reduce damage from severe weather events.

In all history, humans have depended on intuition, shared knowledge, and sensory cues to make successful

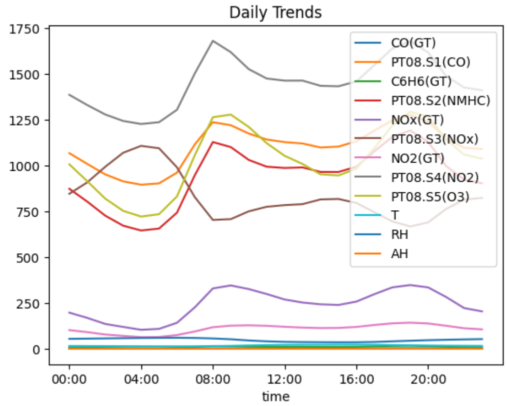

Fig. 2.1: Daily trend

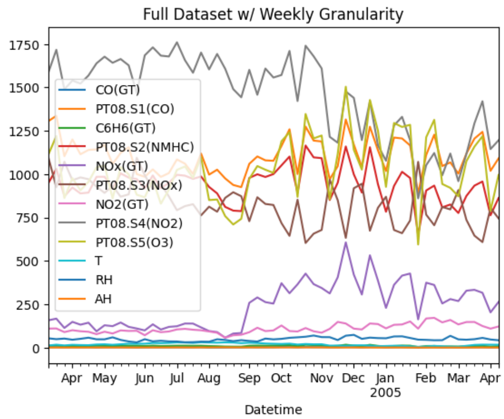

Fig. 2.2: Weekly prediction

decisions in animal husbandry since the early days of domestication. This has significantly improved our practices in animal husbandry and agriculture. The increasing need for food and the progress in sensing technologies could enhance the centralization, size, and efficiency of animal farming. It can revolutionize animal agriculture. This study delves into the challenges and opportunities posed by sensor technology in assisting animal producers to increase meat and animal product production. This study [9]delves at how sensors, big data, artificial intelligence, and machine learning may assist animal producers in reducing production costs, improving efficiency, enhancing animal welfare, and increasing the number of animals per hectare. The text delves into the difficulties and constraints of technology.

On the figure 2.1, the x-axis reflects the passage of time, while the y-axis depicts the level of the substances expressed in parts per billion (ppb). The concentrations of the majority of compounds (CO(GT), C6H6(GT), NOX(GT), and NO2(GT)) tend to change during the course of the day, with values that are typically greater in both the morning and the evening and values that are lower in the afternoon. PT08.S1(CO), PT08.S2(NMHC), PT08.S3(NOx), and PT08.S4(NO2) all exhibit comparable patterns, which suggests that they may be co-located sensors that measure the same air quality conditions. Not only can temperature (T) and relative humidity (RH) fluctuate during the day, but they also have the capacity to affect the quantities of certain compounds.

Figure 2.2 illustrate the timeseries graph, it shows the the daily fluctuations of various air quality measures over the year. These data were most likely gathered at a monitoring station. The date is displayed along the

Table 2.1: Data Sources and Collected Posts

Author	Methods	Contribution	Limitations
Saritha et al [3]	Machine learning	Examines the impact of air pollution on agriculture by analyzing trends in crop production over time	Limited to data availability and the complexity of modelling the relationship between air pollution and crop yields
Elavarasan et al [5]	Support vector regression with radial basis function kernel	Estimates the concentrations of pollutants and the air quality index in California	Limited to the specific case study of California and may not be generalizable to other regions
Albahar et al [6]	Deep learning	Identifies various applications of deep learning in agriculture, including fruit counting, water control, crop management, and disease detection	Highlights challenges such as data complexity, processing resource requirements, and scarcity of deep learning expertise
Aneja et al [7]	Air quality modeling and assessment	Assesses the current status of agricultural air quality research and identifies potential future research avenues	Limited by the uncertainties and complexities involved in air quality modeling
Chen et al [8]	Artificial intelligence (AI)	Examines the use of AI for climate change mitigation and adaptation, including applications in energy efficiency, renewable energy, and sustainable agriculture	Limited by the need for further research and development to address specific challenges and ensure ethical implementation

x-axis, which extends from April of the previous year to March of the current year. A number of different contaminants are represented along the y-axis, and their concentrations are expressed in parts per billion (ppb). As the year progresses, it appears that the concentrations of the majority of pollutants (CO(GT), C6H6(GT), NOX(GT), and NO2(GT)) change. It is possible that the values are greater during the months of October through March as compared to the months of April through September when temperatures are higher. It's possible that this pattern is the result of seasonal shifts in the weather conditions or actions carried out by humans that have an effect on these pollutants. Readings from ground-truth instruments (CO(GT), NOX(GT), and NO2(GT)) and perhaps related sensor readings from particular places (PT08.S1(CO), PT08.S3(NOx), and PT08.S4(NO2)) are included in the measurements for some pollutants, such as carbon monoxide, nitrogen oxides, and nitrogen dioxide. It would appear that the patterns between these values are comparable, which would imply that the sensors are catching circumstances of air quality that are equivalent to one another. In addition, graphs for temperature (T) and relative humidity (RH) are now included in the graph. It is possible for certain environmental conditions to have an effect on the concentrations of some contaminants in the atmosphere.

3. Research Methods.

3.1. Prediction model. Air pollution is the presence of polluting substances that contaminate the air. Air pollution often involves the presence of solid, liquid, and gaseous particles in the outside world. These particles are emitted by fuel and petroleum in automobiles, waste produced from companies or industries in liquid or gas form, ashes produced by volcanoes or wildfires, burning of garbage and fossil fuels, and other sources. Air pollution in real-time is a significant factor in causing chronic illnesses in humans and impacting the environment's natural resources [10]. It also affects the agriculture sector by hindering appropriate crop growth and reducing the productivity of farming. Long-term health issues include nerve damage, lung cancer from inhaling polluted air, kidney failure, and many child health problems. On a global scale, it leads to consequences including ozone layer depletion, acid rain, and global warming, which in turn cause reduced rainfall. A novel ensemble learning model based on a meta-heuristic algorithm is proposed to address the identified issues and improve air quality prediction outcomes.

Generally, the involves incorporating several deep-learning techniques to improve performance. This research utilizes learning techniques such as Bi-LSTM to construct the ensemble model. The methods are executed similarly to a neural network, with neurons capable of categorizing the factors that lead to the outcomes in AQ.

Automated systems are highly proficient at swiftly gathering, handling, and evaluating huge amounts of data. They are unable to make efficient decisions in the absence of data. They can help humans improve decision-making by gathering and analyzing vast amounts of comprehensive data. Various sensors can assist farmers in monitoring animal activities in real-time on a farm. Sophisticated algorithms [11]can utilize large datasets to monitor, measure, and comprehend alterations in animal behavior. Consequently, this can assist farmers in making more informed decisions and implementing timely disease interventions. Air sensors in the poultry sector can now anticipate the beginning of Coccidiosis, an intestinal infection that may rapidly spread among birds without showing any visible signs. One method to detect this illness is by consistently observing air quality. The concentration of volatile organic compounds (VOC) in the air rises with the increasing number of diseased birds. Air sensors can notice this shift earlier than a farmer or doctor. Once the farmers are informed, they can promptly implement measures to halt the illness from spreading. This technique conserves multiple animal lives and averts financial damages.

3.2. Preprocessing. Preprocessing is commonly employed to eliminate redundant features in order to enhance performance. The air data Az is inputted into the first step of the model suggested as a preprocessing strategy. The supplied data typically contain missing values, outliers, and redundant data. Preprocessing is utilized to address these limitations in order to improve the accuracy of the model. Data preparation involves data imputation, data cleansing, and data transformation [12].

Data imputation is utilized to address the absence of values in the input data Az. Missing data are either replaced with a value of zero or estimated using the mean value of the entire sample and the nearest available data point. The data is imputed using an arbitrary data sample represented as A_z^{imp}.

This strategy is employed to identify and eliminate errors and discrepancies. Invalid input. The input data comprises noisy data, outliers, undesirable qualities, and irrelevant data. High computational time occurs when working with irrelevant data, noise, or outliers that lead to errors and inconsistent analysis [13]. Data cleansing is utilized to eliminate redundant data in order to address these issues, resulting in improved performance accuracy and reduced computation time. The resulting data are as follows A_z^{cle}.

Data transformation involves standardizing and consolidating the data. Transforming information is utilized to convert one format of environment data, which includes various sorts of particles such as solid, liquid, and gas, into another one. [14]Transforming the data facilitates predicting air quality and improving performance analysis. The result is what it comes from. The ultimate preprocessed data $A_z^{\mathrm{tra}\,a}$ is then sent on to the feature extraction stage.

3.3. Deep Learning - CNN. The first application of convolutional neural networks is in picture data processing. A multiple-layer perceptron network containing many hidden layers is the structure of a deep network of convolutional neural networks [15][16]. The layer of convolution consists of artificial neurons that represent convolutional filtering and are used to construct feature maps [17]. It is necessary to divide the input into smaller blocks in order to convolve it using a particular set of weights. Through the application of convolutional filters with the same weights to the input, several sets of features may be produced. The pooling layer is applied in order to reduce the number of parameters as well as the dimension of space of the provided data representation. This is accomplished by minimizing the number of parameters. It has been established which is data that is similar in the particular region, because the reaction that is most prevalent is output. An example of a nonlinear function that is utilized with the purpose of learning complex nonlinear structures is the activation functionThe study's learning layers make use of both the exponential and the Rectified Linear Unit (ReLU) [18]. Feature aggregation through global examination of outputs from preceding layers is accomplished by this fully linked learning layer, which is located at the very end of the neural network.

3.4. Long short-term memory. It is the sequence of inputs that determines the output of neural networks with recurrent neurons (RNNs), and the network creates different outputs according to the same input regardless of the order in which the inputs are presented. RNNs combine information from the past with the information that is now being processed during the generation of the output. Long short-term memory, also known as LSTM, is a specialized kind of recurrent neural networks (RNNs) that is employed for the purpose of identifying the long-term dependencies present in sequence data. Data is received externally, stored, recorded in memory cells, and accessed through gates. The memory unit is responsible for controlling the flow of information in order to determine the impact that prior information has on output. Additionally, this unit stores a copy of the predictions that have been made.

After multiplying the weights and information stored in memory, a decision is made on which data will be utilized and how much of it will be used. Some of the weights and information are subsequently added again to the forecast. Some forecasts are chosen as the current prediction, while irrelevant information is isolated to prevent it from influencing future predictions.

The equations for a Long Short-Term Memory (LSTM) cell are as follows:

$$i_t = \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}) \tag{3.1}$$

$$f_t = \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}) \tag{3.2}$$

$$\tilde{C}_t = \tanh(W_{ic}x_t + b_{ic} + W_{hc}h_{t-1} + b_{hc}) \tag{3.3}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{3.4}$$

$$o_t = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}) \tag{3.5}$$

$$h_t = o_t \odot \tanh(C_t) \tag{3.6}$$

Here, x_t is the input at time t, h_t is the hidden state at time t, C_t is the cell state at time t, W and b are weight matrices and bias vectors, σ is the sigmoid activation function, tanh is the hyperbolic tangent activation function, and $\odot$ represents element-wise multiplication.

3.5. Implementation. The initial step of data preparation involved cleaning the data and, in the event that any values were missing for up to five successive time periods, interpolating the data.The minimum-maximum normalization approach was then used to standardize the data.

A set of interpolated time series measurements was constructed using a variety of frame sizes and a number of different data separation techniques.Following the definition of both two-dimensional and three-dimensional input structures,Through the utilization of the Deep Neural Designer Tool that is incorporated into MATLAB edition R2020a, a CNN+LSTM deep learning–based period forecasting framework was constructed. A number of hyper-parameters were adjusted in order to improve the predictive capability to make accurate predictions. Throughout both training and testing, the hyper-parameters were tuned to optimize their performance. It was determined that the structure of the neural network that had the least validating RMSE was the one that should be utilized on test data after the algorithm was executed fifteen times for each approach. Alterations were made to the characteristics and hyperparameters of the neural network, including the hidden layer's type and quantity, the total number of neurons, and the activation function. This allowed for the neural network to be rebuilt and trained.

The Root Mean Square Error (RMSE) is given by:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{3.7}$$

In the following equation:

n is the number of data points or observations.

$\hat{y}_i$ is the predicted value for the i-th observation.

y_i is the true or observed value for the i-th observation.

$\sum_{i=1}^{n}$ denotes the summation over all data points.

Table 3.1 illustrates the comparision of the different model with the evaluation metrices.

Table 3.1: Root Mean Square Error (RMSE) for Different Pollutants

Model	RMSE
LSTM	0.47
BiLSTM	0.25
Proposed Ensemble LSTM+CNN	0.16

3.6. Hyperparameter tuning. Hyperparameters may be associated with the tasks of model selection, such as the topology and size of the network [19], or the pace of learning as well as the size of the mini-batch are two examples of the factors that may be related with the technique for optimizing and the learning procedure among these factors.

This study focused on tuning a specific set of hyperparameters:

Frame Size: The frame size was adjusted using integer values within a specific range.

Step Size: In this investigation, a step size of 1 was used for each time period sample.

Splitting the Data: The training set, validation set, and test set have their data separated by 70%-15%-15% and 80%-10%-10%, respectively.

Selection of Samples: There are three ways to choose a sample: randomly, sequentially, or consecutively. When data samples for training, validation, and testing are chosen at random, this is known as the random selection approach. When data is divided into training and test sets in order, sequential sample selection is used to determine with the use of an evaluation rate, verification samples generated from the original data set. As part of the sequential selection process, the data is broken up into three distinct sections: training, verification, and test procedures.

Validation Frequency: Validation rate is the amount of iterations that occur between assessments of validation metrics. The three values that were taken into consideration for validation frequency were 10, 15, and 20, accordingly. Additionally, it is a reference to the time period during which the validation sample is selected using the ordered choice of sample method.

Sizes of mini-batch: The number of iterations that were selected to be 20, 80, 110, 160, and 200 was done so with the intention of achieving learning progress. This was determined based on the volume of the time series data collection.

Number of epoch: When determining when to stop training the model, this study made use of the early stopping approach.

The number of layers that are convolutional: There were three levels of convolution used in the construction of the CNN part: one, two, and three layers collectively.

Layers of Pooling: When the CNN module was initially developed, there was no pooling layer included in its building process. Following that, a maximum pooling layer and an average pooling layer were applied in the process of building the structure.

Activations Functions: In this particular examination, the Rectified Linear Unit (ReLU) with sigmoid equation were both applied as activation functions simultaneously.

Figure 3.1 illustrated the correlation matrix from an air pollution experiment shows the connections among different pollutants. The correlation coefficient in each cell represents the degree to which two metrics fluctuate together: an amount near 1 indicates a high tendency to rise or decrease in a similar manner, while a value around 0 indicates a weak or missing link. Analyzing these correlations can assist researchers in comprehending the relationships between contaminants and their possible sources or methods of reduction.

CO(GT): Carbon Monoxide (Ground Truth) - perhaps a standard measurement for Carbon Monoxide.

PT08.S1(CO): Probably a Carbon Monoxide sensor reading from a particular position (PT08.S1)

Non-Methane Hydrocarbons (Ground Truth): NMHC(GT)

C6H6(GT): Benzene (Ground Truth)

NOX(GT): Nitrogen Oxides (Ground Truth)

PT08.S2(NMHC): Sensor reading indicating Non-Methane Hydrocarbons

Fig 3.2 shows the a correlation matrix from a study involving partial NMHC (Non-Methane Hydrocarbon)

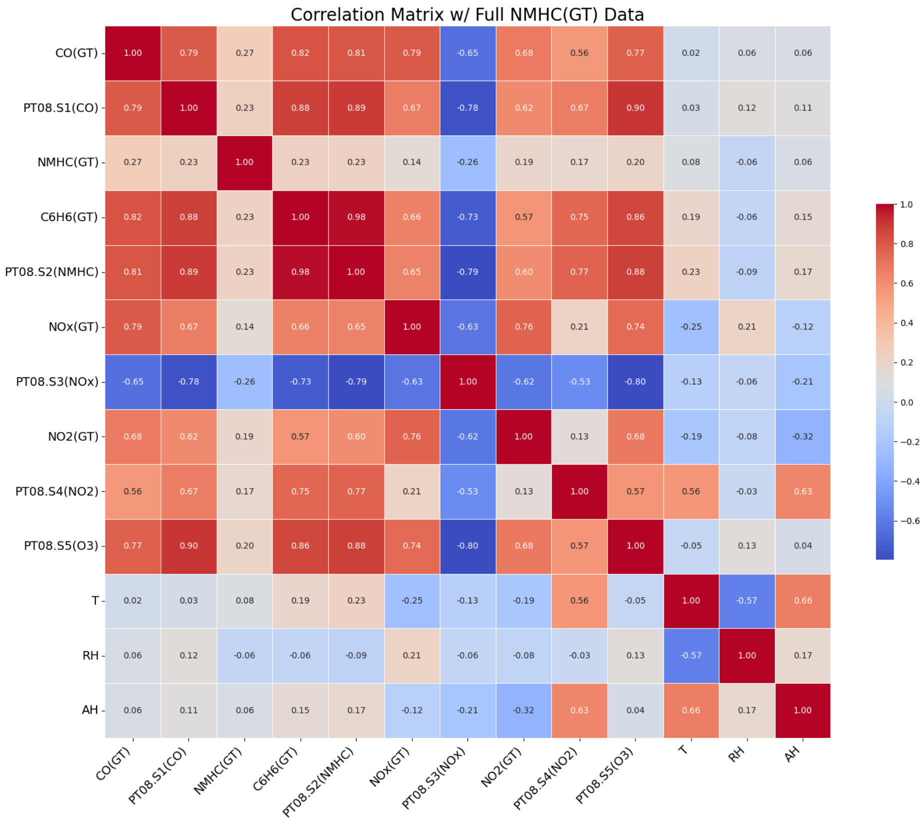

Fig. 3.1: Correlation Matrix

data. The "partial NMHC(GT) data" indicates that the data for non-methane hydrocarbons are inadequate. This might be the result of sensor readings that are either missing or incorrect. Through an analysis of these connections, analysts are able to get an understanding of the ways in which pollutants and other factors may be connected to one another and impact one another.

The heatmap is an illustration in the figure 3.3. The correlation coefficients that exist between air quality indicators and any additional variables that may be significant. A pair of variables, such as carbon monoxide and nitrogen oxides, temperature and humidity, is represented by each individual cell. The degree of correlation and the direction of the correlation are both indicated by the intensity of the color in the cell.

4. Results and Discussions. In pooling layer, which reduces the amount of information that is included in the enormous quantity of input data, the key features of the input are lost, and the size of the input is reduced. In order to make an accurate prediction of the quantity of contaminants using each approach, the model was executed fifteen times for each and every possible arrangement of the hyper parameter [20] variables specified previously. Finally, use the equation 3.7 and the test RMSE as well as the correlation values were computed after the most effective test results were selected based on the values of the RMSE that were the lowest.

During the training phase, the RMSE was found to be at its lowest when the learning rate was set at 0.005. Additionally, the random sample selection approach achieved a lower random sample standard error (RMSE) significance in comparison to the sequential and consecutive techniques of sample selection. As the metric for evaluating the performance of the prediction, the test RMSE values were utilized, and the best possible prediction performance was attained.

Figure 4.1 shoes the scatter plot and show the expected values compared to the actual values of errors,

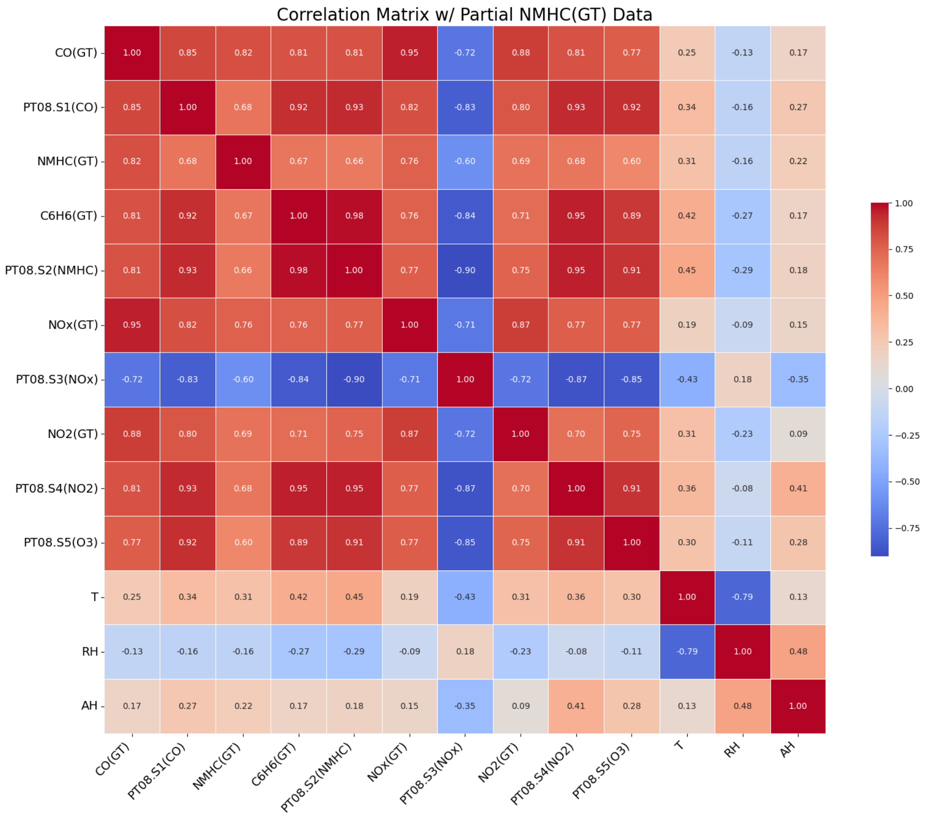

Fig. 3.2: Correlation Matrix

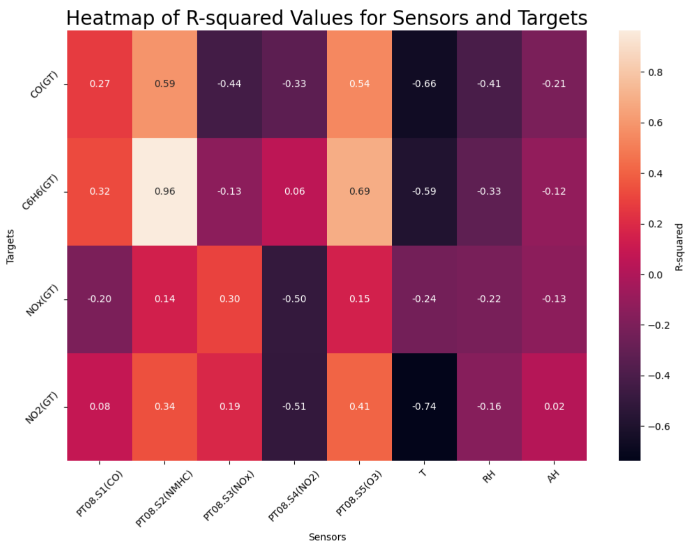

Fig. 3.3: Heat Map

Fig. 4.1: Prediction

as well as the error plotted against time, likely from an air quality study. The x-axis depicts a timeline from March 2004 to March 2005. The left y-axis depicts the error, while the right y-axis reflects the expected and actual values. The graph indicates that the forecasted values consistently exceed the real values, especially in the initial portion of the time frame. The error over time indicates that the forecasted values consistently exceed the real values, with the error diminishing as time progresses.

5. Conclusion. The rapid growth of the industry has led to a concerning problem of air pollution negatively impacting agricultural regions. We have chosen to create an application that predicts the best crop to minimise negative consequences based on existing pollution data. Therefore, the proposed hybrid Deep learning model to predict the air pollutants, and the evaluation metrices Root Mean Square error gives the low error value is 0.16. When constructing the Long Short term Memory model combination with the convolutional neural network in deep learning, we have taken into account the available contaminants. By entering geologically related data, the programme can properly analyse and anticipate the outcome.

REFERENCES

[1] Al-Janabi S, Mohammad M, Al-Sultan A. *A new method for prediction of air pollution based on intelligent computation,Soft Computing [Internet]. 2019 Nov 28;24(1):661–80, 10.1007/s00500-019-04495-1*
[2] Li Y, Sha Z, Tang A, Goulding K, Liu X *The application of machine learning to air pollution research: A bibliometric analysis. Ecotoxicology and Environmental Safety, 2023 Jun;257:114911, dx.doi.org/10.1016/j.ecoenv.2023.114911*
[3] Saritha and Shetty, Ritika and Devi, Manasa and Dhotre, Akash and Hanchinal, Prema R, *Analysis on Air Quality and its Effects on Agriculture*, 2022 3rd International Conference for Emerging Technology (INCET), 2022.
[4] Castelli, Mauro and Clemente, Fabiana Martins and Popovič, Aleš and Silva, Sara and Vanneschi, Leonardo, *A Machine Learning Approach to Predict Air Quality in , Complexity, Hindawi Limited, 2020.*
[5] Elavarasan, Dhivya and Vincent, Durai Raj and Sharma, Vishal and Zomaya, Albert Y. and Srinivasan, Kathiravan, *Forecasting yield by integrating agrarian factors and machine learning models: A survey*
[6] Albahar, Marwan, *A Survey on Deep Learning and Its Impact on Agriculture: Challenges and Opportunities, 13, 2077-0472, Agriculture, MDPI AG, 2023 10.3390/agriculture13030540*
[7] Aneja, Viney P. and Schlesinger, William H. and Erisman, Jan Willem *Effects of Agriculture upon the Air Quality and Climate: Research, Policy, and Regulations Environmental Science Technology, vol. 43, no. 12, pp. 4234–4240, May 2009, doi: 10.1021/es8024403.*
[8] Chen, Lin and Chen, Zhonghao and Zhang, Yubing and Liu, Yunfei and Osman, Ahmed I. and Farghali, Mohamed and Hua, Jianmin and Al-Fatesh, Ahmed and Ihara, Ikko and Rooney, David W. and Yap, Pow-Seng *Artificial intelligence-based solutions for climate change: a review, 2021, 10.1007/s10311-023-01617-y*
[9] *".Deep Learning and Machine Learning based Air Pollution Prediction Model for Smart Environment Design Planning. Global NEST Journal , 2023" n.d. http://dx.doi.org/10.30955/gnj.004735.*
[10] Devasekhar V, Natarajan P. *Prediction of Air Quality and Pollution using Statistical Methods and Machine Learning Techniques. International Journal of Advanced Computer Science and Applications, 2023;14(4), http://dx.doi.org/10.14569/ijacsa.2023.01404103*
[11] Méndez M, Merayo MG, Núñez M. *Machine learning algorithms to forecast air quality: a survey. Artificial Intelligence Review. 2023 Feb 16;56(9):10031–66. http://dx.doi.org/10.1007/s10462-023-10424-4*

[12] Zadmirzaei M, Hasanzadeh F, Susaeta A, Gutiérrez E. *A novel integrated fuzzy DEA–artificial intelligence approach for assessing environmental efficiency and predicting CO2 emissions. Soft Computing. 2023 May 12;28(1):565–91. http://dx.doi.org/10.1007/s00500-023-08300-y*

[13] Sonawani S, Patil K. *Air quality measurement, prediction and warning using transfer learning based IOT system for ambient assisted living. International Journal of Pervasive Computing and Communications. 2023. http://dx.doi.org/10.1007/s00500-023-08300-y*

[14] Singh RB, Patra KC, Pradhan B, Samantra A. *HDTO-DeepAR: A novel hybrid approach to forecast surface water quality indicators. Journal of Environmental Management. 2024. http://dx.doi.org/10.1016/j.jenvman.2024.120091*

[15] Anyaene IH, Onukwuli OD, Babayemi AK, Obiora-Okafo IA, Ezeh EM. *Application of Bio Coagulation–Flocculation and Soft Computing Aids for the Removal of Organic Pollutants in Aquaculture Effluent Discharge. Chemistry Africa. 2023. http://dx.doi.org/10.1007/s42250-023-00754-9.*

[16] Foreback B, Mahura A, Clusius P, Xavier C, Baykara M, Zhou P, et al. *A new implementation of FLEXPART with Enviro-HIRLAM meteorological input, and a case study during a heavy air pollution event. Big Earth Data. http://dx.doi.org/10.1080/20964471.2024.2316320.*

[17] Lakshmipathy M, Prasad MJS, Kodandaramaiah GN. *Advanced ambient air quality prediction through weighted feature selection and improved reptile search ensemble learning. Knowledge and Information Systems. 2023. http://dx.doi.org/10.1007/s10115-023-01947-x*

[18] Janarthanan R, Partheeban P, Somasundaram K, Navin Elamparithi P. *A deep learning approach for prediction of air quality index in a metropolitan city. Sustainable Cities and Society. 2021. http://dx.doi.org/10.1016/j.scs.2021.102720*

[19] Zhang Z, Johansson C, Engardt M, Stafoggia M, Ma X. *Improving 3-day deterministic air pollution forecasts using machine learning algorithms. Atmospheric Chemistry and Physics. 2024. http://dx.doi.org/10.5194/acp-24-807-2024.*

[20] Gholami H, Mohammadifar A, Behrooz RD, Kaskaoutis DG, Li Y, Song Y. *Intrinsic and extrinsic techniques for quantification uncertainty of an interpretable GRU deep learning model used to predict atmospheric total suspended particulates (TSP) in Zabol, Iran during the dusty period of 120-days wind. Environmental Pollution. 2024 http://dx.doi.org/10.1016/j.envpol.2023.123082.*

Edited by: Vadivel Ayyasamy
Special issue on: Internet of Things and Autonomous Unmanned Aerial Vehicle Technologies for Smart Agriculture Research and Practice
Received: Jun 26, 2024
Accepted: Sep 12, 2024