FACTA UNIVERSITATIS
Series: **Electronics and Energetics** Vol. 38, Nº 3, September 2025, pp. 431 - 456
https://doi.org/10.2298/FUEE2503431B

Original scientific paper

HYBRID AI MODEL FOR PREDICTING AIR QUALITY AND RICE CROP YIELD USING SATELLITE DATA AND ENVIRONMENTAL VARIABLES

Dev Ashrit Behera[1], Paras Singh Bhatia[1], Meenakshi Kandpal[1], Jyotirmayee Routray[1], Pranati Mishra[1], Shivani Agarwal[2], Rabindra Kumar Barik[3]

[1]Computer Science and Engineering, Odisha University of Technology and Research, India
[2]Information Technology. Ajay Kumar Garg Engineering College, Ghaziabad, India
[3]School of Computer Applications, KIIT Deemed to be University, India

ORCID iDs:	Dev Ashrit Behera	https://orcid.org/0009-0007-5022-3893
	Paras Singh Bhatia	https://orcid.org/0009-0006-6040-7795
	Meenakshi Kandpal	https://orcid.org/0000-0002-8974-0229
	Jyotirmayee Routray	https://orcid.org/0000-0003-2747-3919
	Pranati Mishra	https://orcid.org/0000-0001-7698-1149
	Shivani Agarwal	https://orcid.org/0000-0002-3233-4449
	Rabindra Kumar Barik	https://orcid.org/0000-0003-3086-3782

Abstract. *Air quality has a significant impact on agricultural productivity, with pollutants such as ozone (O3), particulate matter (PM), and carbon monoxide (CO) posing serious threats to crop yields. Plant growth and yields can be disrupted by these pollutants, highlighting the need for effective solutions. Strategies, including improvements in air quality management to reduce pollutant levels and the development of advanced predictive models, have been proposed. The concentrations of air pollutants and their potential impacts on agriculture can be forecasted by these models, allowing proactive measures to be taken to mitigate adverse effects on crop yields. In this research, the challenges associated with predicting the impact of key air pollutants on rice yield in India are addressed. Satellite data from the Giovanni Data Centre was utilized to monitor concentrations of O_3, PM, and CO, and to calculate the Air Quality Index (AQI). The Prophet model is employed to predict future pollutant levels and AQI. Soil temperature and air moisture data were incorporated to assess their combined impact on crop yield. Nineteen years of monthly rice yield data from FAOSTAT was used to train a feed-forward neural network with inputs including PM, O_3, CO, AQI, soil temperature, and air moisture. A high accuracy of 94% was achieved by the model, effectively predicting crop yields based on these factors, and a clear inverse relationship between air quality and crop yield was demonstrated: significant decreases in yield were correlated with higher concentrations of O_3, PM, and CO.*

Received November 27, 2024; revised January 25, 2025 and March 07, 2025; accepted March 11, 2025
Corresponding author: Dev Ashrit Behera
Computer Science and Engineering, Odisha University of Technology and Research, India
E-mail: devashritbehera@gmail.com

Concrete evidence of the detrimental effects of air pollution on crop productivity is provided by these findings.

Key words: *Air Quality Index, Agricultural productivity, Giovanni Data centre, Prophet Model, Feed Forward Neural Network.*

1. INTRODUCTION

The AQI is a vital tool used globally to measure the quality of the air we breathe and its potential effects on human health and the environment. It assesses the levels of eight primary pollutants in the atmosphere: particulate matter (PM2.5 and PM10), ozone (O3), carbon monoxide (CO), nitrogen dioxide (NO2), sulphur dioxide (SO2), lead (Pb), and ammonia (NH3) [1]. In its most comprehensive form, the AQI integrates multiple pollutant concentrations using a mathematical formula to produce a single AQI value [2].

An AQI value of 100 for each pollutant typically indicates a concentration that matches the temporary national air quality standard for health protection. An AQI below 100 is considered acceptable, while an AQI above 100 signifies hazardous air quality, initially affecting sensitive groups and eventually everyone as the value increases [3]. The National Air Quality Index, launched as part of the Swachh Bharat Abhiyan, is managed by the Central Pollution Control Board with the assistance of State Pollution Control Boards to monitor air quality across numerous cities in India [4]. Despite urbanization, up to 70% of rural households still primarily rely on agriculture for their livelihood [5]. According to the Food and Agriculture Organization of the United Nations, India is the second-largest producer of wheat and rice globally [6]. Crop yield, defined as the amount of crop produced per unit area, is a crucial metric in agricultural productivity. Several factors influence crop yield, including soil quality, irrigation practices, climate conditions, and increasingly, environmental pollutants. As noted in several studies [7,8,9], air pollutants have a significant impact on crop yields.

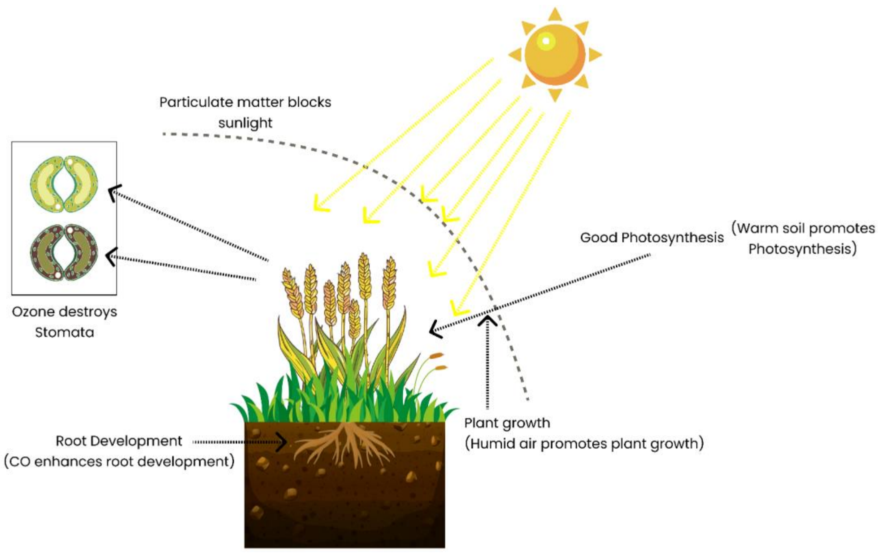


Fig. 1 Various environmental effects on crops

Figure 1 illustrates the effect of air pollutants, along with air temperature and soil moisture, on crops. Ozone (O3) is an environmental gaseous pollutant that enters leaves through stomatal pores, causing foliage damage [10]. Particulate matter accumulation on leaves can reduce photosynthesis [11]. Although less researched in agricultural contexts, carbon monoxide can disrupt plant respiration processes [12]. India's diverse climatic zones and varied agricultural practices make it particularly vulnerable to the effects of air pollution. The Green Revolution, a transformative period in Indian agriculture, aimed to alleviate hunger and food insecurity through advanced technologies [13]. While it transformed India into an agricultural powerhouse, it also led to increased pesticide and insecticide use, contributing to environmental degradation. Understanding the interplay between air quality and agricultural productivity is essential for developing strategies to mitigate the adverse effects of pollution on crops. This context highlights the importance of ongoing research and policy initiatives aimed at improving air quality and safeguarding agricultural productivity.

1.1. Motivation

Our motivation is inspired by the novelty and interdisciplinary nature of our approach, which combines geospatial data analysis, AI modelling, and agricultural science to address a critical issue: the impact of air quality on crop productivity. Traditional methods of assessing crop yield have often overlooked the intricate and dynamic interplay between environmental pollutants and agricultural outcomes. By leveraging advanced technologies such as satellite data from Giovanni and statistical data from FAOSTAT, alongside sophisticated AI techniques, our approach aims to fill this gap. This innovative integration not only enhances our understanding of how pollutants like O3, PM, and CO influence rice yield but also provides actionable insights for mitigating environmental risks and optimizing agricultural practices in a changing world. Our work is driven by the potential to contribute to sustainable agriculture and global food security through cutting-edge environmental informatics.

1.2. Objectives

The research aims to determine the relationship between air pollutants, AQI, temperature, moisture, and rice production using satellite data analysis, artificial intelligence (AI) modeling, and agricultural science methodologies. Specifically, our objectives include assessing the concentration of gases, including O_3, PM, and CO, using satellite data from Giovanni [14] and calculating the AQI to understand the environmental conditions affecting agricultural regions; using the Prophet model to predict future trends in air pollutants and AQI levels; analyzing the impact of air pollutants, temperature, and moisture concentrations on rice yield utilizing data from the Food and Agriculture Organization of the United Nations [15], aiming to understand the complex interplay between environmental factors and agricultural outcomes; and developing a feedforward neural network model to analyze the relationship between air pollutants, AQI, temperature, moisture, and crop yield to identify underlying patterns and key drivers of crop productivity.

1.3. Organizations

The structure of this paper is as follows: Section 2 summarizes relevant studies, offering an overview of methods and findings related to air quality indices (AQI), rice yield, and crop yield predictions. Section 3 presents the proposed framework, detailing the methodology, data sources, and predictive models used in the study. Section 4 discusses the implementation and results, covering the algorithms employed, the training process and the accuracy of our predictions. Finally, Section 5 concludes by highlighting the main findings and contributions, and proposing directions for future research.

2. RELATED WORK

Extensive research has been conducted in recent years to understand the impact of air pollutants on crop yields and to develop accurate predictive models for agricultural productivity. Many studies have explored the detrimental effects of pollutants like ozone (O3), particulate matter (PM), and carbon monoxide (CO) on crop growth. In addition, advanced modelling techniques, such as deep learning and time series forecasting, have been employed to enhance prediction accuracy. This section reviews key contributions in these areas.

L. Zhoul et al. [16] focused on air pollution in Nanjing by analysing the AQI using data from June 2014 to May 2019. They employed the Prophet model to predict AQI levels from June 2019 to April 2020, providing insights into future air quality trends in the region. Similarly, Sama et al. [17] addressed global air pollution concerns, emphasizing the impact of fossil fuels and vehicular emissions. Their research demonstrated that time series forecasting was more effective than linear regression for predicting pollution levels, particularly in time-dependent datasets. They applied their model to air quality data from Bhubaneswar, India, yielding strong predictive accuracy within significant confidence intervals. Setianingrum et al. [18] explored air quality issues in Jakarta, one of the most polluted cities in the world. They utilized the Prophet model and various forecasting techniques to predict pollutants like PM10, SO2, CO, and O3 using ISPU DKI Jakarta data from 2010 to 2019. Their findings indicated that the Prophet model was more effective than ARIMA for predicting SO2, CO, and O3, while ARIMA performed better for PM10 and NO2. Further extending the application of the Prophet model, Shen et al. [19] used it to predict air pollution levels in South Korea. They optimized the model's parameters using three years of air quality data from Seoul Open Data Plaza and successfully predicted concentrations of PM2.5, PM10, SO2, and CO up to one year in advance.

In the field of agricultural productivity, Ji et al. [20] developed models using historical yield data and weather variables, adjusting artificial neural network (ANN) parameters to optimize prediction accuracy. Their study demonstrated that ANN models outperformed multiple linear regression models in predicting rice yields, achieving higher R^2 values and lower RMSE values under Fujian's climatic conditions. Building on this work, Chu Z et al. [21] introduced the BBI model, which integrates neural networks to forecast rice yields in Guangxi Zhuang Autonomous Region. This approach achieved the highest accuracy for both summer and winter rice yield predictions, illustrating the potential of neural networks in agricultural forecasting. Dahikar et al. [22] explored the impact of climatological phenomena on crop production, focusing on the effectiveness of artificial neural networks (ANNs) for prediction. Their research highlighted methodologies for predicting crop yields based on soil and atmospheric parameters, further advancing the application of machine learning in agriculture. Nevavuori et al. [23] applied Convolutional Neural Networks (CNNs) to predict crop yields using data from UAVs, specifically NDVI and RGB data. They optimized CNN performance through hyperparameter tuning and demonstrated that RGB data provided more accurate yield predictions compared to NDVI data, showcasing the potential of UAV-based remote sensing in precision agriculture.

Turning to air pollution's impact on agriculture, S. Pandya et al. [24] reviewed research on the effects of air pollution on agricultural productivity, particularly in developing countries like India. Their study focused on yield losses due to air pollution, using AQI metrics to represent variations in air quality across different regions and highlighting the recent impacts of particulate matter on agricultural towns and regions. H. Jethva et al. [25] analyzed satellite imagery and agricultural data to study the correlation between post-monsoon rice production, vegetation index, agricultural fires, and air pollution in northwestern India. Their research found a significant increase in crop production, vegetation index, and agricultural fire activity, leading to heightened aerosol loading and PM2.5 levels over the Indo-Gangetic Plain. The study emphasized the need for effective crop residue management to mitigate hazardous air quality in the region. Ghosh et al. [26] focused on improving rice grain yield predictions in Bhubaneswar, India, by using forecast data from a wide-range forecasting system. They downscaled forecast data to daily weather sequences and incorporated it into a crop simulation model. Their findings demonstrated improved prediction accuracy as the season progressed, aiding decision-making for rice farming. Future research could refine these techniques and extend the application to other crops and regions. Dhekale et al. [27] tested the reliability of the Extended Range Forecasts System (ERFS) for predicting Kharif rice yields in Kharagpur, West Bengal. They used the CERES-Rice model and stochastic weather generators to convert ERFS forecasts into daily sequences for input into crop models. Their results showed that ERFS forecasts effectively predicted year-to-year variability in rice yields, helping farmers make informed decisions for climate risk management in rice production.

Finally, S. Mishra et al. [28] explored the impact of seasonal changes on rice yield along the Odisha coast in India using machine learning techniques. They proposed three models: a classifier ensemble, regression techniques with boosting, and a feature ranking and fusion method. Their models provided valuable insights into the climatic effects on rice yield, offering promising solutions distinct from traditional prediction methods. Turning to the role of IoT in fostering smart environments, V. Terzieva et al. [29] explored the transformative potential of IoT technologies in various contexts, with a particular focus on smart schools and education. The study highlighted how IoT devices could optimize learning environments by enhancing the efficiency of the educational process. Two prototype IoT devices were implemented as part of the research, featuring a communication protocol designed to collect diverse sensor data and provide critical insights using laser technology. The experimental outcomes demonstrated significant promise, with plans underway for developing advanced prototypes. The study underscores the potential of IoT in shaping smart educational spaces, paving the way for innovative learning methodologies and resource management [35].

Table 1 Notable contributions to AQI, rice yield, and crop yield predictions

Author	Model Used	Findings
L.Zhoul et al. [16]	Prophet, Prophet-SVR, Prophet-LSTM	Enhanced prediction accuracy of AQI using hybrid models in Nanjing.
Samal et al. [17]	SARIMA, Prophet	Demonstrated the effectiveness of time series forecasting models over linear regression for global air pollution.
Setianingrum et al. [18]	Prophet, ARIMA	Showed the impact of the Prophet model in air quality prediction in Indonesia (Jakarta).
Shen et al. [19]	Prophet	Successfully predicted six pollutants and extended prediction time in Seoul, South Korea.
Ji et al. [20]	ANN	Demonstrated superior rice yield predictions in Fujian's climatic conditions over linear regression models.
Chu Z, Yu J et al. [21]	BBI-model, BPNNs, IndRNN	Accurate forecasts of rice yields in Guangxi Zhuang Autonomous Region, China, for both summer and winter seasons.
Dahikar et al. [22]	ANN	Discussed the effectiveness of ANN in predicting crop yield based on climatological phenomena.
Nevavuori et al. [23]	CNN	Achieved lower errors in crop yield prediction using UAV RGB data compared to NDVI.
S.Pandya et al. [24]	Sarima, Prophet	Reviewed the impact of air pollution on agriculture in developing countries like India.
H.Jethva et al. [25]	Normalized Difference Vegetation Index (NDVI)	Highlighted the need for effective crop residue management based on the correlation between crop production, vegetation index, agricultural fires, and air pollution.

2.1. Research gap

While significant progress has been made in understanding the relationship between air quality and crop yield, several critical gaps remain unaddressed. Many studies focus on individual pollutants or use simpler linear regression models, which fail to capture the intricate, non-linear interactions between multiple environmental factors and crop yield. For example, the works by Ji et al. [20] and Chu et al. [21] primarily rely on historical yield data and basic weather variables, without incorporating the synergistic effects of pollutants like CO, PM, and O3 with soil temperature and air moisture. Similarly, studies using machine learning techniques often do not leverage advanced time-series forecasting models to predict future environmental conditions, limiting their ability to provide actionable insights for proactive agricultural management. Another limitation in the current body of research is the lack of integration between satellite-derived pollutant data and machine learning-based predictive modeling.

Although some works utilize geospatial data, they often fail to preprocess it comprehensively or address its temporal dynamics, reducing prediction accuracy. Furthermore, studies exploring the application of the Prophet model are typically confined to air quality forecasting without linking these predictions to their downstream impacts on agricultural productivity. To address these gaps, our study combines the time-series forecasting capabilities of the Prophet model with the predictive power of a Feed Forward Neural Network. This novel integration allows for the simultaneous forecasting of air pollutants and the modeling of their direct and indirect impacts on crop yield. Additionally, by incorporating geospatial satellite data and multi-year FAOSTAT yield data, our approach provides a comprehensive, data-driven analysis that bridges the divide between environmental science and agricultural informatics.

3. PROPOSED FRAMEWORK

3.1. Methodology

In agricultural forecasting, integrating diverse data sources and advanced algorithms significantly enhances prediction accuracy. The proposed model employs the Prophet framework to analyse monthly CO, PM, and O3 concentrations from 2005 to 2023, sourced from Giovanni data archives. By incorporating trend, seasonality, and additional regressors, the model captures long-term trends, periodic patterns, and external influences on pollutant concentrations, which are then used to determine the AQI. This AQI, along with predictions of soil moisture and air temperature, is crucial for accurate crop yield forecasting. The training data, combined with crop yield data from the same period, ensures comprehensive and reliable predictions for improved decision-making and resource management in agriculture.

Figure 2 describes the entire model consisting of data inflow from Giovanni data archives and FAOSTAT data centres. It describes the accumulation of a dataset that is to be loaded onto the Feed Forward Neural Network involving processed data dimensions of the pollutant concentration and AQI. It also shows the creation of an input dataset which is to be prompted for crop yield prediction by taking predicted pollutant concentration from the Prophet model along with the AQI from the predicted data. It clearly shows the AQI generation process as well. Finally, crop yield was predicted from the input data and results were shown.

Following successful model training, crop yield prediction is executed utilizing the input dataset generated by the preceding model. This involves employing sophisticated time series forecasting techniques and rigorous AQI calculations to facilitate accurate predictions. The integration of these datasets allows for a comprehensive analysis of the intricate interplay between environmental factors and crop yield outcomes.

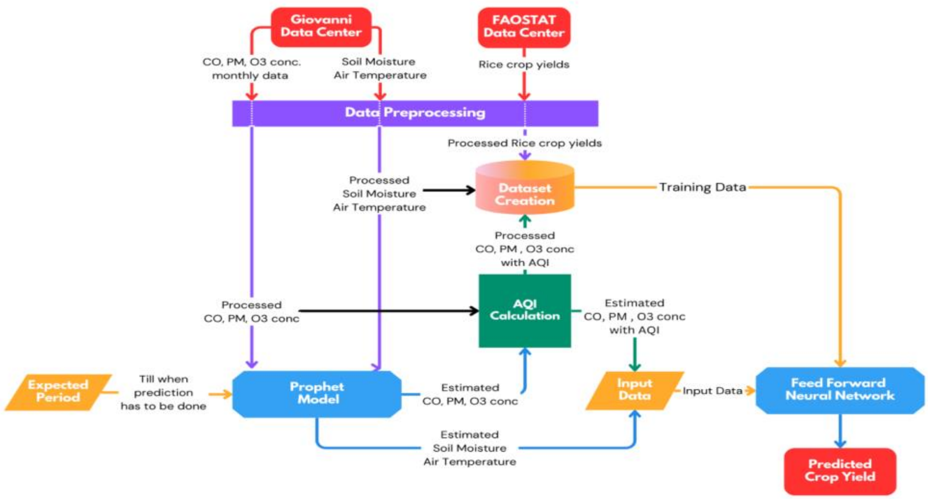


Fig. 2 Schematic Diagram of the AQI Prediction and Crop Yield Analysis Model

This holistic approach ensures a robust and reliable framework for forecasting agricultural productivity, thereby offering valuable insights for informed decision-making in the realm of food security and agricultural policy.

3.1.1. Hypothesis for method selection

The selection of the Feed Forward Neural Network (FNN) and the Prophet model is rooted in their complementary capabilities, making them well-suited for the complex task of predicting crop yields based on air quality. The Prophet model is specifically designed for time series forecasting and excels in handling missing data, capturing seasonality, and identifying long-term trends. Its ability to forecast pollutant concentrations such as CO, PM, and O3, along with the Air Quality Index (AQI), provides a robust foundation for understanding future environmental conditions. These predictive insights are crucial as they establish the baseline data required for accurate crop yield predictions. On the other hand, the FNN is a powerful tool for modeling non-linear relationships in high-dimensional datasets. It is particularly effective in processing multidimensional inputs, including pollutant concentrations, AQI, soil temperature, and air moisture, to predict crop yields. The neural network’s architecture allows it to learn complex patterns and dependencies, providing precise results even in scenarios where relationships between variables are intricate and not immediately apparent. By combining these two methods, we leverage the strengths of both statistical forecasting and machine learning. The Prophet model complements the FNN by serving as an accurate preprocessor, generating reliable predictions for environmental variables. These predictions are integrated into the FNN as input features, allowing the neural network to explore how these variables interact with one another and influence agricultural productivity. This synergy enhances the overall accuracy and interpretability of the predictions, addressing both temporal trends and complex dependencies in the data.

3.2. Dataset Description

3.2.1. Giovanni Data Centre

It is a web application created by NASA. The first requirement for the successful prediction of crop yield at a particular period in the future requires the AQI, soil moisture and air temperature at that time which is determined by the Prophet model. The raw concentration of three major contributing pollutant gases is taken into account for the determination of AQI. The concentration of carbon monoxide (CO) refers to the number of CO molecules within an atmospheric column extending from the Earth's surface up to the stratosphere per square centimetre of surface area. Dust (particulate matter) consists of small solid particles that can either remain suspended in the atmosphere as aerosols or accumulate as sediment on the Earth's surface. In the troposphere, ozone (O_3) forms naturally from diatomic oxygen (O_2) through electric discharge in air or ultraviolet radiation, with urban areas exhibiting higher ozone concentrations due to anthropogenic pollution.

MERRA-2 [30] is NASA's latest global atmospheric reanalysis dataset, spanning from 1980 to the present, produced by the Global Modelling and Assimilation Office using their working model, typically updated approximately every three weeks after each month's end. All the gases come from different collections of this dataset. The other two factors affect the crop yield growth directly. Average layer soil moisture refers to the depth-averaged water content within a specific soil layer below the Earth's surface, while the warmth of the air indicates the movement energy in the atmosphere at a specific location.

Table 2 presents an analysis of data sourced from various dimensions, including pollutants and environmental factors obtained from Giovanni. Each dimension is associated with a specific collection, unit of measurement, and frequency of data collection.

Table 2 Analysis of environmental pollutants and factors from Giovanni

Dimension	Collection	Obtained Unit	Frequency
Carbon Monoxide	M2TMNXCHM	ppbv	Monthly
Particulate Matter	M2TMNXAER	$kg\ m^{-3}$	Monthly
Ozone	M2IMNXASM	dobsons	Monthly
Soil Moisture	M2TMNXLND	$m^3\ m^{-3}$	Monthly
Air Temperature	M2IMNXLFO	K	Monthly

3.2.2. FAOSTAT Data centre

FAOSTAT is an online database maintained by FAO. FAOSTAT is a critical resource for policymakers, researchers, and the general public, offering data on various aspects of agriculture. Some FAOSTAT Domains Overview [31] are mentioned in the next lines. Production covers agricultural production data, including quantities produced, producer prices, value at the farm gate, harvested area, and yield per hectare. Trade provides annual trade statistics for about 600 food and agriculture commodities since 1961, collected from national authorities and international organizations. Food Security includes food supply data, crucial for global and national undernourishment assessments and economic analysis. Prices contain annual and monthly producer prices, producer price indices, and consumer price indices. Resources include data on the national distribution of land (arable, pasture, and other lands) and the importance of irrigation. Investment contains data on private investment in agriculture, official development assistance, and government spending.

3.3. Data Preprocessing

The transformation of raw data into a clean, consistent, and well-defined format, which can be utilized for modelling and analysis, is generally aimed at data preprocessing. Based on the source of data, various approaches to preprocessing need to be carried out to ensure the data's accuracy, reliability, and suitability for analytical tasks.

3.3.1. Giovanni Data

The geospatial data coming from Giovanni data archives was pre-processed by three processes. Filtering concentration data and collecting it based on a single location across each month, followed by performing data cleaning on the collected data. The concentration is then converted into suitable units to ensure accuracy while calculating AQI. Finally, the concentration is normalized before forecasting to enhance predictive accuracy. In the Earth coordinate system, the latitudes to the north of the equator are taken positive and the south of the equator are taken negative, whereas longitudes to the east of Prime Meridian are taken positive and west of Prime Meridian are taken negative. The dataset covers -180° to 180° longitude and -90° to 90° latitude, with a resolution of 0.5° x 0.625°. The top-left corner is (-180°, -90°), and the bottom-right is (180°, 90°). The analysis focuses on a single location's coordinates, using latitude and longitude as inputs. Data values at that location are filtered for the period from 2005 to 2023. Missing values, except for soil moisture (NaN), had large negative placeholders, which were replaced with the mean of legitimate values. NaN values were replaced with zero. Gas concentrations were converted to appropriate units for AQI calculation. Table 3 provides a Conversion Table for Gaseous Concentration, specifying the calculated unit, along with the corresponding multiplying factor for each dimension.

Table 3 Gaseous concentration conversion table with their current dimension and required dimension

Dimension	Calculated Unit	Multiplying factor
CO	mg m^{-3}	1.15 times e^{-3}
PM	µ m^{-3}	e^{9}
O3	Dobsons	0.1

The dimensions of soil moisture and air temperature did not require conversion as they were not involved in the calculation of AQI but rather served as important dependencies of crop yield and growth. At last, scaling the concentration to normalize the data to a range between 0 and 1 was done using min-max scaling. This normalization prevented some concentrations from dominating the model's training process simply because they have larger magnitudes.

3.3.2. FAOSTAT data

This study utilizes a dataset on rice yield spanning 19 years, from 2005 to 2023, sourced from the Food and Agriculture Organization Corporate Statistical Database (FAOSTAT). FAOSTAT, managed by the Food and Agriculture Organization of the United Nations (FAO), is a globally recognized repository for agricultural data, providing reliable and comprehensive statistics. The dataset includes the following columns: Domain Code, Domain, Area Code (M49), Area, Element Code, Element, Item Code (CPC), Item, Year Code, Year, Unit, Value

For this study, the dataset was refined to retain only the most relevant columns, including year, yield values, and associated attributes, streamlining the analysis. The dataset is publicly accessible for reproducibility and further research at the FAOSTAT website: https://www.fao.org/faostat. Researchers are encouraged to use this dataset for replicating experiments or conducting additional studies. Figure 3 Depicts the dataset from the FAOSTAT having the column names

Domain Code	Domain	Area Code (M49)	Area	Element Code	Element	Item Code (CPC)	Item	Year Code	Year	Unit	Value
QCL	Crops and livestock products	356	India	5412	Yield	0113	Rice	2019	2019	kg/ha	4083.6
QCL	Crops and livestock products	356	India	5412	Yield	0113	Rice	2020	2020	kg/ha	4075.7
QCL	Crops and livestock products	356	India	5412	Yield	0113	Rice	2021	2021	kg/ha	4196.2
QCL	Crops and livestock products	356	India	5412	Yield	0113	Rice	2022	2022	kg/ha	4257.1
QCL	Crops and livestock	356	India	5412	Yield	0113	Rice	2023	2023	kg/ha	4322.3

Fig. 3 Raw Data from FAOSTAT with a yield value of 100g/ha

3.4. AQI Determination

At the heart of AQI computation lie several primary pollutants, including particulate matter, ozone, and carbon monoxide. These pollutants, originating from diverse sources such as vehicle emissions, industrial processes, and natural events, have distinct chemical compositions and health impacts. Particulate matter consists of tiny particles suspended in the air, while ozone and nitrogen dioxide are reactive gases formed through atmospheric reactions. Table 4 likely represents the AQI and what it signifies about the quality of air. It typically categorizes air quality into different levels, providing information about the future health risks belonging to it.

Table 4 AQI representing health implications over a range

AQI	Intensity	Description
0-50	Good	The air quality is acceptable, though sensitive individuals may have some health concerns.
51-100	Moderate	The air quality is within acceptable limits, but sensitive individuals may still have some health concerns.
101-150	Mildly Unhealthy	Air quality is deemed unhealthy for vulnerable groups such as children, the elderly, and those with respiratory or heart conditions.
151-200	Unhealthy	The air quality has reached unhealthy levels, and adverse health effects may be experienced by everyone.
201-300	Very Unhealthy	The air quality has deteriorated to a very unhealthy level, with potential significant health effects for the entire population.
301-500	Hazardous	Air quality is hazardous, and everyone is at risk of experiencing more serious health effects.

As a result, there is a safe limit of exposure for every pollutant contributing to AQI. Table 5 describes the standards for the determination of AQI as per US. Environmental Protection Agency. It describes the concentration of pollutants in the air up to which the air quality can be considered acceptable.

Table 5 AQI and concentration of pollutants

AQI	PM10 ($\mu g\ m^{-3}$)	CO ($mg\ m^{-3}$)	O3 ($\mu g\ m^{-3}$)
0-50	0-50	0.0-1.0	0-50
51-100	51-100	1.1-2.0	51-100
101-200	101-250	2.1-10	101-168
201-300	251-350	10.1-17	169-208
301-400	351-430	17.1-34	209-748
401-500	430+	34+	748+

The formula for the calculation of AQI is mentioned below [32], that is

$$I_p = \frac{(I_{High} - I_{Low})(C_p - BP_{low})}{(BP_{High} - BP_{Low})} + I_{Low} \tag{1}$$

where I_p= Index value for pollutant p, C_p= Measured concentration of pollutant p, BP_{High}= The nearest concentration breakpoint that is greater than or equal to C_p, BP_{Low}= The nearest concentration breakpoint that is less than or equal to C_p, I_{High} = AQI value associated with BP_{High} and I_{Low} = AQI value associated with BP_{Low}.

3.5. Dataset Creation

This section describes the creation of the dataset that is used for training the Feedforward Neural Network (FNN) model. This comprehensive dataset integrates environmental parameters and agricultural productivity metrics, ensuring a robust input for our predictive analysis.

3.5.1. Preprocessed Environmental Data Collection

We sourced environmental data from the Giovanni data center, which included the following parameters: ozone, particulate matter, carbon monoxide, Air Moisture and Soil Temperature.

3.5.2. AQI Calculation

Figure 4 shows the process of integrating the calculated AQI with the environmental data. Using the concentrations of O3, PM, and CO, we calculated the AQI. The AQI provides a standardized metric that reflects the combined effects of various pollutants on air quality. This calculation was crucial for creating a single, comprehensive measure of air pollution. The calculated AQI was then merged with the environmental data, which includes air moisture and soil temperature.

	pm_conc	co_conc	o3_conc	aqi	moisture_conc	temperature_conc
2005-01-01	15.687535	0.194340	25.393068	26.145892	0.186029	294.293091
2005-02-01	20.314533	0.177672	25.164618	33.857555	0.150872	298.814941
2005-03-01	18.073730	0.154295	26.767704	30.122884	0.134153	302.124420
2005-04-01	47.634571	0.129381	27.696100	79.106689	0.112684	304.326355
2005-05-01	45.404739	0.124172	28.217482	75.339042	0.139401	306.058380

Fig. 4 Merging of Environmental Data And AQI

3.5.3. Rice Yield Data

Rice yield data for India, spanning from 2005 to 2019, was obtained from the FAOSTAT database. This data provided annual rice yield values, offering a historical perspective on agricultural productivity.

3.5.4. Merging Datasets

Figure 5 illustrates the process of creating a unified dataset for training the Feed Forward Neural Network (FNN) model. The steps include data alignment by ensuring temporal alignment of environmental data and calculated AQI with corresponding rice yield values, and combining features by merging all relevant environmental parameters

(O3, PM, CO, air moisture, soil temperature, and calculated AQI) with rice yield data. This comprehensive dataset forms a robust foundation for training the FNN model to learn the critical relationships between air quality and crops.

	Unnamed: 0	pm_conc	co_conc	o3_conc	aqi	moisture_conc	temperature_conc	Value
0	2005-01-01	15.687535	0.194340	25.393068	26.145892	0.186029	294.29310	0.614977
1	2005-02-01	20.314533	0.177672	25.164618	33.857555	0.150872	298.81494	0.554836
2	2005-03-01	18.073730	0.154295	26.767704	30.122884	0.134153	302.12442	0.539812
3	2005-04-01	47.634570	0.129381	27.696100	79.106689	0.112684	304.32635	0.398747
4	2005-05-01	45.404740	0.124172	28.217482	75.339042	0.139401	306.05838	0.359672

Fig. 5 Final Dataset For the Input to Feed Forward Neural Network

3.5.5. Final Dataset Composition

The final dataset included the columns: O3 (Ozone concentration), PM (Particulate matter concentration), CO (Carbon monoxide concentration), AQI (Calculated Air Quality Index), Air Moisture (Air moisture levels), Soil Temperature (Soil temperature levels), and Rice Yield (Annual rice yield values from FAOSTAT).

Table 6 Summary of three datasets combined together

Dataset	Source	Key Attributes
Environmental Data	Giovanni Data Center	Concentrations of O3, PM, CO; soil temperature; air moisture. Data is time-series (monthly) and spans 2005–2023.
Air Quality Index	Derived Calculation	Computed from environmental data using the AQI formula. Reflects combined pollutant levels as a single metric.
Rice Yield Data	FAOSTAT	Historical rice yield values (annual) in India, spanning 2005–2023. Serves as the dependent variable for predictions.

3.6. Prophet Model

The statistical approach of examining a sequence of data points over some time to look out for trends and patterns is called Time Series Forecasting. It is used to understand the underlying structure and function of the data, which aids in making forecasts and informed decisions. By decomposing the series into its fundamental components—trend, seasonality, and noise—analysts can better interpret the data and predict future values. Trend refers to the overall change of effect in the long term in which minor effects are neglected. Seasonality refers to regular and predictable changes that recur at specific intervals. Noise refers to the insignificant fluctuation which needs to be eliminated from data points The Prophet Model is a decomposable additive model based on the generalized additive model architecture. Its components work together to produce results by combining linear and nonlinear trends with seasonalities and the effects of holidays. This combination ensures accurate and reliable forecasting, making it particularly effective for time series with missing data or outliers.

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{2}$$

where g(t) is the growth model capturing the trend (e.g., logistic or piecewise linear), s(t) is the seasonality component capturing periodic patterns (e.g., daily, weekly, yearly), h(t) represents the holiday effects, modeled with dummy variables for special events and ϵ_t is the noise or residual error, accounting for unexplained variance.

Figure 6 depicts the basic architecture of the Prophet Model. It comprises a growth model responsible for capturing overall trends, which intersects with the seasonality component. Finally, it incorporates holiday effects and provides the resultant trend, coupled with noise residuals or errors.

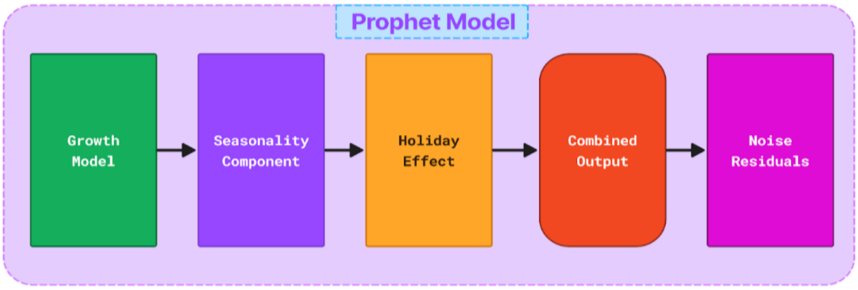


Fig. 6 Underlying structure of Prophet Model

3.6.1. Growth Model

The first component of the Prophet model is the growth model, which captures underlying trends. It offers two options: the Logistic Growth Model, suitable for data with saturating, non-linear growth, accounting for carrying capacity and growth rate, and useful when data plateaus after rapid growth; and the Piecewise Linear Model, ideal for linear data with clear growth or shrinkage trends, involving a constant growth rate. Both models can include change points, where the growth rate changes, either specified manually or selected automatically, ensuring accurate forecasting of various growth patterns.

$$g(t) = k + \sum_{j=1}^{m} a_j (t - b) l_{t-b} \tag{3}$$

where k is the base growth rate, b are the change points and a_j are adjustments to the growth rate at change points

3.6.2. Seasonality Component

The seasonality component in the Prophet model captures recurring patterns or variations at fixed intervals within time series data. This is crucial for forecasting scenarios influenced by factors like weather, holidays, or cultural events. The seasonality component includes several subcomponents, each representing different periodicities such as daily, weekly, monthly, or yearly. These subcomponents combine to form the overall seasonality effect. By accurately capturing seasonality, the Prophet model produces more reliable and precise forecasts, allowing users to make informed decisions based on underlying patterns in the data.

$$s(t) = \sum_{n=1}^{N} \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right) \tag{4}$$

where P is the period (e.g., 365.25 for yearly seasonality), N is the number of Fourier terms to include and a_n and b_n are coefficients learned by the model.

3.6.3. Holiday effect

The holiday effect in the Prophet model adds dummy variables to capture the impact of holidays and special events on time series data. Analysts provide a list of past and future events, and the model assigns parameters for changes in the forecast during those times. Extra parameters account for effects over surrounding days, ensuring accurate forecasting by incorporating holiday impacts into the model.

$$h(t) = \sum_{l=1}^{L} a_l l_{t \in H} \tag{5}$$

where H is the set of periods associated with holiday l and a_l is the parameter indicating the effect size for holiday l.

3.6.4. Noise Residuals

The error term or residuals in a model capture unexplained variance, reflecting its accuracy. Analysts adjust model parameters to minimize these errors, improving forecast reliability. By identifying and addressing large errors, outliers, or sudden increases, analysts refine their models. Reviewing flagged forecasts helps make necessary adjustments, ultimately enhancing prediction accuracy.

$$\in_t \sim N(0, \sigma^2) \tag{6}$$

where σ^2 is the variance of the residuals.

3.7. Feed Forward Neural Network

A Feedforward Neural Network is a basic type of artificial neural network where data flows in a single direction—from input, through hidden layers, to output. Neurons in FNNs process inputs with weighted sums and activation functions, forming the basis for more complex neural network designs. Despite their simplicity, FNNs are versatile and widely used in machine learning applications.

3.7.1. Layers of FNN

In a Feed Forward Neural Network (FNN), the input layer receives initial data, with each neuron representing a feature of the input vector. It distributes this data to hidden layers for processing. Hidden layers perform complex transformations using weighted connections and activation functions, allowing the network to learn intricate patterns. The number and configuration of these layers and neurons can be adjusted based on task complexity. The output layer produces final predictions or classifications; its neurons correspond to output classes or regression targets. In classification tasks, softmax activation converts outputs into class probabilities. A loss function, such as cross-entropy for classification, compares predicted outputs to true labels, optimizing model performance by minimizing errors and adjusting predictions.

Figure 7 depicts a feedforward neural network with layers (including input and output). The input layer comprises n nodes representing the input vector X. Each subsequent layer (l = 1 to L-1) contains hidden neurons $w_{ij}w_{ij}$ denotes the weight connecting the neuron in the layer to the neuron in the layer l. The final layer has output neurons (Y). A softmax function converts these outputs into a probability distribution over classes. The loss function measures the difference between the network's prediction (Y) and the true labels (Z). Backpropagation utilizes this loss to iteratively adjust the weights ($w_{ij}w_{ij}$) to minimize the overall error, enabling the network to learn complex relationships within the data.

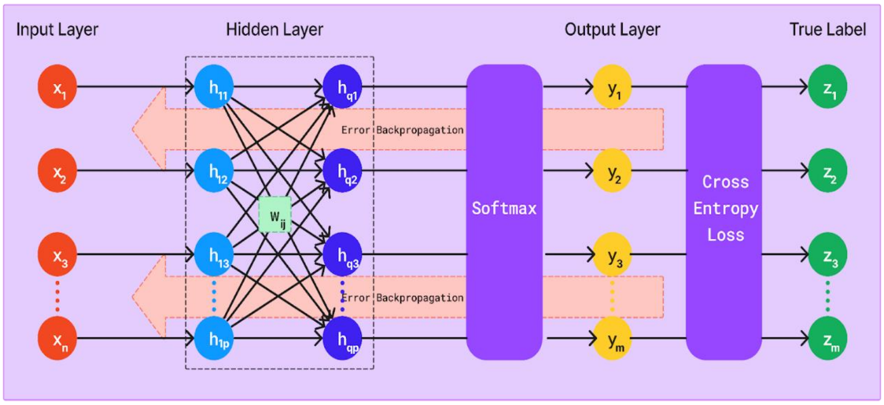


Fig. 7 Underlying structure of Feed Forward Neural Network

3.7.2. Softmax function

In neural networks designed for classification tasks, the softmax function in the output layer converts raw output scores (logits) into a probability distribution over classes. It ensures outputs range from 0 to 1 and sum to 1, crucial for interpreting them as probabilities. The function emphasizes the most probable class by assigning it a higher probability while reducing probabilities for less likely classes.

3.7.3. Error Backpropagation

Error backpropagation is essential for neural network learning, calculating gradients of the loss function with respect to each weight to minimize prediction errors. It efficiently updates weights by propagating error gradients backward from the output layer to the input layer using the chain rule, ensuring optimal adjustments for minimizing loss across the network.

3.7.4. Entropy Loss Function

The cross-entropy loss function assesses classification model performance by measuring the disparity between predicted and true probability distributions. It penalizes incorrect classifications more severely, facilitating effective neural network training. As predicted probabilities deviate from actual labels, cross-entropy loss increases, guiding the network towards more accurate predictions.

3.7.5. Model Performance Evaluation

To evaluate the performance of the Feed Forward Neural Network (FNN), two key metrics were computed: Mean Squared Error (MSE) and a derived Accuracy score. The MSE measures the average squared differences between the predicted values and the true values, giving an indication of the model's prediction error. A lower MSE suggests better model performance, as the predictions are closer to the actual values. In this model, the MSE was calculated on the test dataset to quantify the prediction error. To provide a more intuitive performance measure, the MSE was also converted into an "accuracy" score. This approach gives us a relative measure of model performance, where a lower MSE leads to a higher accuracy score. Both metrics offer valuable insight into the model's effectiveness, with MSE reflecting the error and the accuracy score offering a user-friendly metric for overall model performance. Figure 8 Shows The accuracy and MSE of the FNN Model depicting the precision of the model.

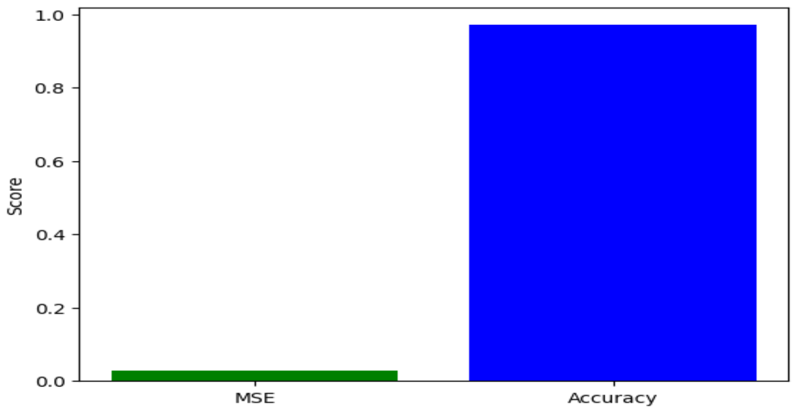


Fig. 8 Model Performance Analysis – MSE and Accuracy Metrics Mean Squared Error (MSE): 0.029, Converted Accuracy: 0.971

4. RESULT AND DISCUSSION

The data frame used with the Prophet Model typically includes two columns: 'ds' for timestamps in YYYY-MM-DD format and 'y' for numerical values, such as normalized pollutant concentrations. By providing a desired future period, like five years converted to months, to the predict() function, the model forecasts future values based on historical data. The data frame returned by the Prophet model after successful forecasting contains several predicted components, which can be utilized for comprehensive analysis. These consist of yhat, yhat_upper, and yhat_lower, where yhat represents the main forecasted value for each date, while yhat_upper and yhat_lower provide the uncertainty intervals as upper and lower bounds, respectively. The trend, trend_upper, and trend_lower represent the estimated trend component of the time series, showing values without considering seasonal changes, with trend_upper and trend_lower indicating the bounds of uncertainty in the trend. Additionally, yearly, yearly_upper, and yearly_lower capture seasonality on a yearly scale, with yearly_upper and yearly_lower representing the uncertainty bounds for this component.

Table 7 Configuration of Prophet Model

Parameter	Description	Value/Setting
Growth Model	Captures underlying trends in the data. Two options: logistic or linear.	Linear
Seasonality	Handles periodic changes (daily, weekly, yearly).	Yearly
Change Points	Points where trend changes significantly	Automatically detected
Holiday Effects	Includes the impact of holidays or special events on forecast	Not used
Uncertainty Interval	Range within which the values are expected to fall	95%
Forecasting Period	Period for which predictions are made	5 years (60 months)
Data Input Columns	Two required columns: ds (timestamps) and y (normalized pollutant concentrations).	As specified in the dataset creation section

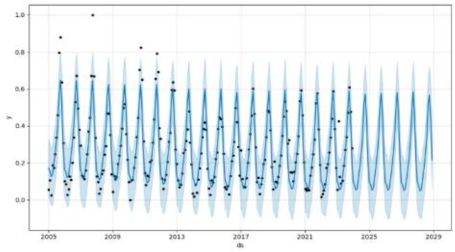
Figure 9 shows the DataFrame returned after successfully forecasting carbon monoxide gas levels using the Prophet model.

	ds	yhat	yhat_lower	yhat_upper	trend	yearly	yearly_lower	yearly_upper
0	2005-01-01	0.170878	0.018669	0.331257	0.314967	-0.144089	-0.144089	-0.144089
1	2005-02-01	0.156013	0.011364	0.313959	0.314473	-0.158460	-0.158460	-0.158460
2	2005-03-01	0.125193	-0.032761	0.282807	0.314027	-0.188835	-0.188835	-0.188835
3	2005-04-01	0.133625	-0.018072	0.285023	0.313533	-0.179909	-0.179909	-0.179909
4	2005-05-01	0.188827	0.033903	0.340458	0.313056	-0.124229	-0.124229	-0.124229

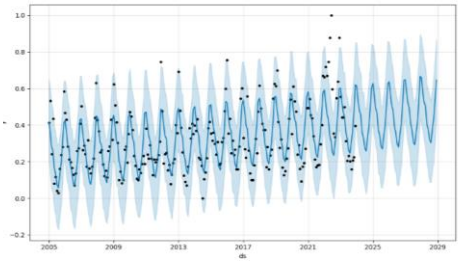
Fig. 9 Major components of forecasted DataFrame for CO pollutant

After achieving the predicted data frame for all the taken pollutants namely CO, PM and O3, the *plot()* function was used to generate a visualization of the forecasted values along with the observed data. The visuals consisted of predefined data which showed the actual data points plotted earlier, forecasted values which meant the predicted values that were represented on the plot and uncertainty intervals which around the forecasted values represent the range within which the true values are expected to fall with a specified probability.

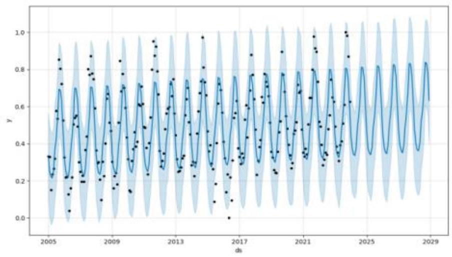
Figure 10 plots generated using the plot () function depict various pollutants with yhat representing normalized concentration values on the y-axis. Original data points are marked by black dots, while a dark blue line shows predicted values (yhat), accompanied by a light blue shaded region indicating uncertainty (yhat_upper and yhat_lower bounds). The graphs display monthly data up to 2023, followed by predictions. Carbon monoxide (Figure 10(a)) initially decreases and stabilizes, despite scattered original data. Particulate matter (Figure 10(b)) shows closer data points and a steepening trend in predicted values. Ozone (Figure 10(c)) exhibits widely distributed data points and an observable increase in predicted concentrations over the years.



(a)



(b)

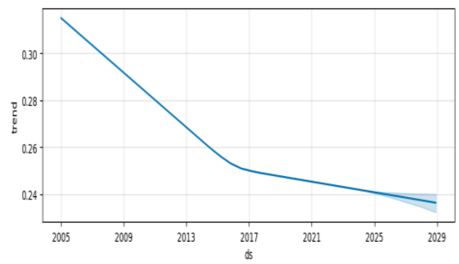


(c)

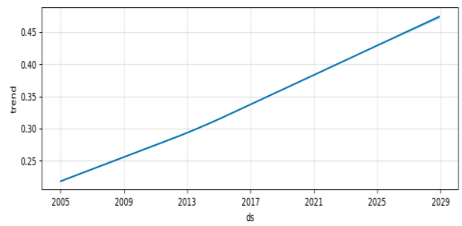
Fig. 10 (a) Represents the plot of the carbon monoxide pollutant, (b) Represents the plot particulate matter pollutant, (c) Represents the plot of ozone pollutant

The trend component graph visualizes the overall trend in the data. On the x-axis, it displays the years, providing a timeline over which the data spans. This allows for a clear view of how the trend evolves over time. The y-axis represents the trend values, reflecting the underlying direction and magnitude of the data after seasonal effects have been removed. This component is essential for identifying long-term patterns, such as whether the data shows a general increase, decrease, or remains stable.

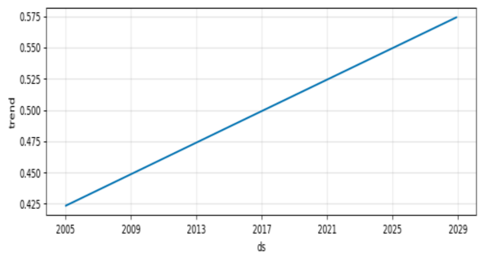
Figure 10 illustrates the graphs of the trend component for all pollutants over the years from 2005 to 2023. It shows almost negligible uncertainty, suggesting a high level of confidence in the direction and rate of change indicated by the trend. Figure 11(a) describes the trend for carbon monoxide, which seems to decrease greatly up to 2017, after which the slope of the trend decreases with a little uncertainty towards the end. Figure 11(b) describes the trend for particulate matter, where it linearly increases except for a small increase in slope around 2014. The values of the trend almost lie between 0.2 to 0.5. Figure 11(c) describes the trend for ozone, where it simply shows an increasing trend; however, the slope is less compared to particulate matter.



(a)



(b)

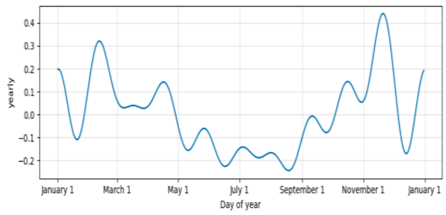


(c)

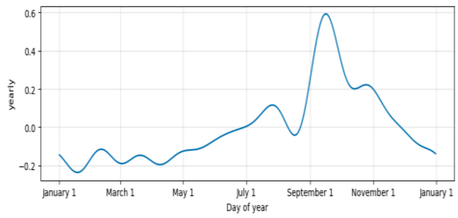
Fig. 11 (a) Visualization of trend component of CO (b) Visualization of trend component of PM (c) Visualization of trend component of O3

The yearly component graph visualizes the annual variations within the data. Along the x-axis, it presents the days throughout the year, delineating the timeline across which the data extends. This offers a comprehensive perspective on the yearly fluctuations over time. The y-axis depicts the yearly component values, encapsulating the recurring patterns and deviations from the overall trend across each year, after accounting for seasonal effects. This component proves invaluable for discerning cyclic trends, such as annual peaks or troughs, and for understanding how the data behaves within each yearly cycle.

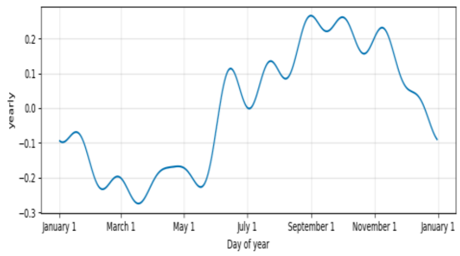
Figure 12 shows yearly seasonality graphs for pollutants, highlighting cyclic seasonal variations. The non-linear yearly components reflect these cyclic patterns within the data, occasionally resulting in negative values indicating lower observed values compared to expected trends. Figure 12(a) illustrates CO's seasonality with a peak in late September and consistently low values throughout the year. Figure 12(b) shows PM's seasonality, peaking in late November with a secondary peak in February. Figure 12(c) displays O3's seasonality, peaking in September and maintaining high values from July to January, with lower values in other months.



(a)



(b)



(c)

Fig. 12 (a) Visualization for a yearly component of CO (b) Visualization for a yearly component of PM (c) Visualization for trend component of O3

The "yhat" component contains all the predicted values, but they are initially in normalized form. To obtain the original values, we utilize the "min-max scaler" inverse_transform() function, which converts the values back to their original scale. Then, we calculate the AQI using the method described previously. Additionally, considering the significance of factors such as soil moisture and air temperature for crop growth and yield, we incorporate the predicted values of these variables for the corresponding dates. This enriched dataset is then structured into a DataFrame, combining all relevant dimensions. Finally, this comprehensive DataFrame serves as input for predicting crop yield using a trained Feed Forward neural network. The performance of the Feed Forward Neural Network model was tested using test datasets. Figure 13 shows each test dataset includes calculated AQI values along with pollutant concentrations (PM2.5, O3, CO), air moisture, and soil temperature. These datasets were fed into the trained FNN to predict the corresponding rice yield values. The results highlight a clear relationship between air quality and crop yield.

	Unnamed: 0	pm_conc	co_conc	o3_conc	aqi	moisture_conc	temperature_conc
0	01-01-2025	44.541105	0.076497	27.938566	73.879799	0.108921	304.716837
1	01-04-2025	47.541105	0.096497	29.938566	25.979893	0.110080	310.716837
2	01-07-2025	40.541105	0.196497	25.938566	40.872853	0.210080	319.716837

Fig. 13 Test Dataset for Feed Forward Neural Network

The FNN is configured based on the optimizations mentioned in table 8 to give the most accurate and precise results which were found to be reasonable and logical.

The results from the FNN model, shown in Figure 14, clearly indicate that an increase in AQI, which signifies deteriorating air quality, is associated with a decrease in rice yield. This inverse relationship underscores the significant impact of air pollution on agricultural productivity. Specifically, Case 1 had the highest AQI, reflecting the poorest air quality among the three cases. Consequently, the predicted rice yield was the lowest, illustrating the detrimental effects of severe pollution on crop productivity.

Table 8 Configuration of the Feed Forward Neural Network

Component	Description	Value/Setting
Input Layer	Receives processed inputs: pollutant concentrations, AQI, soil temperature, air moisture, etc.	6 nodes
Hidden Layers	Layers between input and output layers to capture non-linear relationships.	2 layers, 10 neurons each
Activation Function	Non-linear transformation applied to hidden layer outputs.	ReLU (Rectified Linear Unit)
Output Layer	Produces the final prediction (rice yield).	1 node (for regression output)
Loss Function	Measures the difference between predicted and true values.	Mean Squared Error (MSE)
Optimization Algorithm	Adjusts weights to minimize the loss function.	Adam optimizer
Learning Rate	Controls the step size of updates in the optimization algorithm.	0.001
Epochs	Number of times the entire dataset is passed through the network during training.	100
Batch Size	Number of samples processed before updating the model weights.	32

Case 2 shows that with the lowest AQI, indicating the best air quality, this dataset showed the highest predicted rice yield. This result underscores the benefits of clean air for agricultural output. Case 3 exhibited moderate AQI levels, resulting in a moderate predicted rice yield. The results from this case further validate the correlation between air quality and crop yield, with moderate pollution leading to correspondingly moderate yields.

```
AQI: 73.87979893 - Predicted Crop Yield: 0.9687761068344116
AQI: 25.97989322 - Predicted Crop Yield: 0.9990944862365723
AQI: 40.8728528 - Predicted Crop Yield: 0.9896674156188965
```

Fig. 14 Relation Between AQI and Crop Yield

4.1. Ablation Study

To evaluate the effectiveness of combining datasets, we compared models trained on individual datasets (FAOSTAT and Giovanni) with those using the hybrid dataset. The FAOSTAT dataset, while essential for providing historical yield data, lacks environmental metrics, limiting its predictive accuracy. Giovanni data, with its comprehensive environmental factors, improves model performance but does not address direct agricultural impacts. The hybrid dataset combines these strengths, leading to better overall results. Similar findings by Wijayanti et al. [33] emphasize the importance of integrating diverse datasets to capture complex dependencies in agricultural forecasting.

Table 9 Dataset Comparison Results

Dataset	MSE	MAE	R^2
Only FAOSTAT	0.041	0.162	0.78
Only Giovanni	0.029	0.139	0.84
Hybrid Dataset	0.018	0.102	0.94

To further analyze the improvements brought by combining methods, we compared baseline models (Prophet and FNN) with the hybrid model. Prophet alone performs well in forecasting temporal trends but struggles with non-linear interactions, while FNN effectively models such complexities but lacks temporal forecasting capabilities. Combining these methods leverages their individual strengths, resulting in significant performance gains. These findings align with Setiadi et al. [34], who demonstrated the efficacy of hybrid approaches in integrating temporal and non-linear modeling techniques.

Lastly, we compared our hybrid model with popular methods from related works. ARIMA, while effective for linear trends, shows limitations in handling complex, non-linear interactions. LSTM improves upon ARIMA by addressing non-linearities but lacks dataset integration capabilities. The hybrid model, while slightly less accurate than XGBoost, provides a balanced approach with fewer computational requirements and enhanced dataset synergy, making it more adaptable for broader applications.

Table 10 Comparison with Related Work

Method	Dataset	MSE	MAE	R^2
ARIMA	FAO & WB	0.045	0.150	0.78
XGBoost	FAO & WB	0.012	0.089	0.96
LSTM	Regional Dataset	0.025	0.130	0.86
Prophet Only (Baseline)	Giovanni Dataset	0.032	0.145	0.82
FNN Only (Baseline)	FAOSTAT	0.026	0.133	0.85
Hybrid Model	Hybrid Dataset	0.018	0.102	0.94

The code for this paper is available at https://zenodo.org/records/14203376

5. Conclusion and Future Work

In conclusion, utilizing the Feed Forward Neural Network (FNN) model with input derived from the forecasted values of the Prophet model, we observed a significant inverse relationship between Air Quality Index (AQI) and crop yield. This relationship highlights how increasing AQI, which signifies worsening air quality, adversely impacts agricultural productivity. These findings underscore the critical importance of considering environmental factors like air quality when analysing and improving agricultural outcomes. Poor air quality can directly or indirectly affect plant growth by reducing photosynthetic activity, altering the chemical composition of soil and water, and impacting the overall ecosystem health. Such insights are essential not only for agricultural planning but also for implementing effective environmental policies to mitigate the negative impacts of pollution.

Future studies could further enhance this research by incorporating additional factors, such as soil quality, precipitation levels, temperature variations, and pest outbreaks, to provide a comprehensive understanding of the interplay between environmental conditions and crop yield. Leveraging advanced modeling techniques, including cutting-edge machine learning methodologies and deep learning architectures, can significantly improve predictive accuracy, reliability, and the ability to capture complex nonlinear relationships among variables. Extending the scope to include diverse geographic regions and a variety of crops will make the findings more generalizable and actionable for global agricultural strategies. Additionally, developing real-time monitoring systems for data collection and integration with predictive models can aid in dynamic decision-making, helping farmers adapt quickly to changing environmental conditions. By addressing these aspects, future research can contribute to the development of robust solutions for optimizing agricultural productivity in the face of environmental challenges. This work also underscores the need for interdisciplinary collaboration between environmental scientists, agricultural researchers, and data scientists to tackle the multifaceted issues of food security and environmental sustainability.

References

[1] R. Huang, R. Hu and H. Chen, "A Novel Hybrid Model for Air Quality Prediction via Dimension Reduction and Error Correction Techniques", *Environ. Monit. Assess.*, vol.197, p. 96, 2025.

[2] M. Suman, "Air Quality Indices: A Review of Methods to Interpret Air Quality Status", *Mater. Today*, vol. 34, no. 3, pp.863-868, 2020.

[3] Kanchan, A. K. Gorai and P. Goyal, "A Review on Air Quality Indexing System", *Asian J. Atmos. Environ.*, vol. 9, pp. 101-113, 2015.
[4] P. K. Meena and D. V. Singh, "Air Quality Monitoring and Pollution Control Technologies", *Int. J. Multidiscip. Res. Sci. Eng. Technol.*, vol. 7, pp.4409-4426, 2024.
[5] L. S. Pereira, "Water, Agriculture and Food: Challenges and Issues", *Water Resour. Manag.*, vol. 31, 2017, pp. 2985-2999.
[6] A. Singh and A. Gupta, "Food Security Through Sustainable Agriculture: A Prospective Study in the Indian Context", In *Sustainability and Health Informatics: A Systems Approach to Address the Climate Action Induced Global Challenge*, Singapore: Springer Nature, 2024, pp. 155-182.
[7] H. Kaur, "Air Pollution and Greenhouse Gases Emissions: Implications in Food Production and Food Security", in *Greenhouse Gases: Sources, Sinks and Mitigation*, Singapore: Springer, 2022, pp. 107-133.
[8] D. Tong, R. Mathur, K. Schere, D. Kang and S. Yu, "The Use of Air Quality Forecasts to Assess Impacts of Air Pollution on Crops: Methodology and Case Study", *Atmos. Environ.*, vol. 41, pp. 8772-8784, 2007.
[9] W. Wei and Z. Wang, "Impact of Industrial Air Pollution on Agricultural Production", *Atmosphere*, vol. 12, pp. 1-15, 2021.
[10] M. M. Hasan, M. A. Rahman, M. Skalicky, N. M. Alabdallah, M. Waseem, M. S. Jahan, G. J., Ahammed, M. M. El-Mogy, A. A. El-Yazied, M. F. Ibrahim and X. W. Fang, "Ozone Induced Stomatal Regulations, MAPK and Phytohormone Signaling in Plants", *Int. J. Mol. Sci.*, vol. 22. pp. 12, 2021.
[11] J. K. Lee, D. Y. Kim, S. H. Park, S. Y. Woo, H. Nie and S. H. Kim, "Particulate Matter (PM) Adsorption and Leaf Characteristics of Ornamental Sweet Potato (Ipomoea batatas L.) Cultivars and Two Common Indoor Plants (Hedera helix L. and Epipremnum aureum Lindl. & Andre)", *Horticulturae*, vol. 8, p. 1, 2022.
[12] M. Wang and W. Liao, "Carbon Monoxide as a Signaling Molecule in Plants", *Front. Plant Sci.*, vol. 7, p. 572, 2016.
[13] M. A. Chakwanda, P. T. Chenge, N. Baloyi, G. L. Tamburayi, M. Muchangana, P. Jarbah, A. Rera and J. P. Sahoo, "Green Revolution: The Catalyst for Agricultural Transformation", *Vigyan Varta: Int. J. E-Magazine Sci. Enthus.*, vol. 5, pp. 294-302, 2024.
[14] Giovanni Satellite Data Center, NASA. [Online] Available at: https://www.earthdata.nasa.gov/technology/giovanni
[15] World Health Organization, "Asia and the Pacific Regional Overview of Food Security and Nutrition 2020: Maternal and Child Diets at the Heart of Improving Nutrition", *Food & Agriculture Org.*, vol. 2020, 2021.
[16] L. Zhoul, M. Chenl and Q. Ni, "A Hybrid Prophet-LSTM Model for Prediction of Air Quality Index", In Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence, 2020, pp. 595-601.
[17] K. K. R. Samal, K. S. Babu, S. K. Das and A. Acharaya, "Time Series Based Air Pollution Forecasting Using SARIMA and Prophet Model", In Proceedings of the International Conference on Information Technology And Computer Communications, 2019, pp. 80-85.
[18] A. H. Setianingrum, N. Anggraini and M. F. D. Ikram, "Prophet Model Performance Analysis for Jakarta Air Quality Forecasting", In Proceedings of the 10th International Conference on Cyber and IT Service Management (CITSM), 2022, pp. 1-7.
[19] J. Shen, D. Valagolam and S. McCalla, "Prophet Forecasting Model: A Machine Learning Approach to Predict the Concentration of Air Pollutants (PM2. 5, PM10, O3, NO2, SO2, CO) in Seoul, South Korea", *Peer J.*, vol. 8, p. e9961, 2020.
[20] B. Ji, Y. Sun, S. Yang and J. Wan, "Artificial Neural Networks for Rice Yield Prediction in Mountainous Regions", *J. Agric. Sci.*, vol. 145, pp. 249-261, 2007.
[21] Chu, Zheng, and Jiong Yu. "An End-To-End Model for Rice Yield Prediction Using Deep Learning Fusion", *Comput. Electron. Agric.*, vol. 174, p. 105471, 2020.
[22] S. S. Dahikar and S. V. Rode, "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach", *Int. J. Innov. Res. Electr. Electron. Instrum. Control Eng.*, vol. 2, pp. 683-686, 2014.
[23] P. Nevavuori, N. Narra and T. Lipping, "Crop Yield Prediction with Deep Convolutional Neural Networks", *Comput. Electron. Agric.*, vol. 163, p. 104859, 2019.
[24] S. Pandya, T. R. Gadekallu, P. K. R. Maddikunta and R. Sharma, "A Study of the Impacts of Air Pollution on the Agricultural Community and Yield Crops (Indian Context)", *Sustainability*, vol. 14, p. 13098, 2022.
[25] H. Jethva, O. Torres, R. D. Field, A. Lyapustin, R. Gautam and V. Kayetha, "Connecting Crop Productivity, Residue Fires, and Air Quality Over Northern India", *Sci. Rep.*, vol. 9, p. 16594, 2019.
[26] K. Ghosh, A. Singh, U. C. Mohanty, N. Acharya, R. K. Pal, K. K. Singh and S. Pasupalak, "Development of a Rice Yield Prediction System Over Bhubaneswar, India: Combination Of Extended Range Forecast and CERES-Rice Model", *Meteorol. Appl.*, vol. 22, pp. 525-533, 2015.

[27] B. S. Dhekale, M. M. Nageswararao, A. Nair, U. C. Mohanty, D. K. Swain, K. K. Singh and T. Arunbabu, "Prediction of Kharif Rice Yield at Kharagpur Using Disaggregated Extended Range Rainfall Forecasts", *Theor. Appl. Climatol.*, vol. 133, pp. 1075-1091, 2018.

[28] S. Mishra, "Prediction of Climatic Variability Effects on Rice Yield of Coastal Odisha using Machine Learning Technique, Odisha", PhD Thesis, 2020.

[29] V. Terzieva, S. Ilchev and K. Todorova, "The Role of Internet of Things in Smart Education.", *IFAC-Papers On Line*, vol. 55, pp. 108-113, 2022.

[30] R. Gelaro, W. McCarty, M. J. Suárez, R. Todling, A. Molod, L. Takacs, C. A. Randles, A. Darmenov, M. G. Bosilovich, R. Reichle and K. Wargan, "The Modern-Era Retrospective Analysis for Research and Applications", *J. Clim.*, vol. 30, pp. 5419-5454, 2017.

[31] FAOSTAT, FAO. "Food And Agriculture Organization Corporate Statistical Database." FAO Online Database, 2022. [Online] Available at: http://www. fao. org/faostat/en/# data

[32] S. A. Horn and P. K. Dasgupta, "The Air Quality Index (AQI) in Historical and Analytical Perspective a Tutorial Review", *Talanta*, vol. 267, pp. 1-53, 2024.

[33] E. B. Wijayanti, D. R. I. M. Setiadi and B. H. Setyoko, "Dataset Analysis and Feature Characteristics to Predict Rice Production based on eXtreme Gradient Boosting", *J. Comput. Theor. Appl.*, vol. 1, no. 3, pp. 299-310, 2024.

[34] D. R. I. M. Setiadi, A. Susanto, K. Nugroho, A. R. Muslikh, A. A. Ojugo and H.-S. Gan, "Rice Yield Forecasting Using Hybrid Quantum Deep Learning Model", *Computers*, vol. 13, no. 8, p. 191, 2024.

[35] M. Djordjević, B. Jovičić, S. Marković, V. Paunović and D. Danković, "A Smart Data Logger System Based on Sensor and Internet of Things Technology as Part of the Smart Faculty", *Journal of Ambient Intelligence and Smart Environments*, vol. 12, pp. 359-373, 2020.