A DEEP LEARNING MODEL FOR AIR QUALITY PREDICTION

Presented by
ABHIMANYU A S (LTRV20IT069)
GATHA REGHUNATH (TRV20IT033)
SAYANDH K S(TRV20IT058)
MERLYN DEENA PHILIPOSE(TRV20IT062)

Under the Guidance of
Prof.Jassim Rafeek

Dept. of Information Technology
Government Engineering College Barton Hill

May 10, 2024

Outline

Introduction

- Monitoring and predicting air quality levels help protect public health by giving advance notice of potential health risks from bad air quality.
- People, especially vulnerable groups like kids, older adults, and those with respiratory issues, can take precautions to stay safe.
- Real-time alerts and notifications about variations in air quality will help us to take necessary health precautions to an extent.

Literature Survey

Title	Source	Description
Comprehensive Deep Learning Model for Air Quality Prediction in Smart Cities	IEEE International Journal on Big Data, 2022	Presents a deep learning model tailored for air quality prediction in smart cities. Utilizes advanced neural network architectures to analyze complex environmental data and forecast air quality levels with high accuracy. Integrates deep learning techniques with IoT sensor data for real-time air quality monitoring and prediction.

Title	Source	Description
Development of an Air Quality Prediction System	IEEE North Karnataka Subsection Flagship International Conference, 2020	This conference paper delves into the development of an air quality prediction system that incorporates both machine learning (ML) and deep learning (DL) techniques. The study underscores the efficacy of DL models in capturing intricate patterns within air quality data, surpassing the performance of conventional ML approaches. By harnessing DL algorithms, the proposed system achieves enhanced accuracy and reliability in forecasting air quality levels.

Title	Source	Description
Air Pollution Prediction Using Machine Learning	IRJET Volume 10, 2022	The paper explores air pollution prediction through machine learning methodologies. One notable aspect emphasized is the significance of hardware connections utilized for data collection. This inclusion underscores the critical role of hardware in ensuring the accuracy and reliability of collected data, which serves as the foundation for training machine learning models. By effectively integrating hardware sensors and devices into the data collection process, the study aims to enhance the quality of input data, consequently improving the predictive capabilities of the machine learning algorithms employed for air pollution forecasting.

Title	Source	Description
Text to Speech Conversion	Indian Journal of Science and Technology, October 2016	The paper proposes the utilization of the gtts Python module for text-to-speech conversion. Specifically, text-to-speech conversion is employed to convert textual information, including the peak PM2.5 value, date, and health precautions, into speech format. This approach facilitates the auditory communication of crucial information, enhancing accessibility and usability for individuals who may benefit from audio-based content consumption.

Title	Source	Description
Python Based Gesture Control Using PyAutoGUI	JETIR April 2023, Volume 10, Issue 4	This paper introduces the utilization of the PyAutoGUI module in Python for gesture control. The PyAutoGUI module enables automation of gestures within the system. Specifically, the authors employed it for automating the process of sending WhatsApp broadcast alerts. This approach enhances efficiency and ease of use by automating repetitive tasks, thereby streamlining the process of sending alerts through gestures.

Problem Statement

- Detecting and predicting air quality is crucial for keeping people and controlling air pollution.
- According to studies, Particulate matter within the diameter of 2.5 micrometers or smaller are major air pollutants which are harmful for health.
- The challenge involves measuring and predicting future PM 2.5 values based on historical and current data.
- Early warning plays an important role in health precautions.

Specific Objectives

- To utilize IoT data to measure values of ozone and nitrogen dioxide ,Carbon monoxide and Particulate matter(PM) levels.
- To develop a robust deep learning model to analyze IoT data to forecast future levels of PM 2.5 using LSTM networks.
- To develop a Graphical User Interface(GUI) showing the predicted PM2.5 graph within the selected date ranges.

Specific Objectives(Contd..)

- To develop a method for finding peak PM 2.5 value within the predicted range of date and to convert it into an audio file with peak value,date and health advisory.
- To develop a subprocess for the automation of broadcasting of this audio file through WhatsApp.

Proposed Solution

- The most important gases causing air pollution are ozone , carbon monoxide, nitrogen dioxide and particulate matter.
- These gases can be measured using IOT sensors.
- LSTM algorithm can be used for the time series prediction of PM2.5 value which can be done using keras and tensorflow and a specific function is created for finding peak PM2.5 value .

Proposed Solution(Contd..)

- Higher PM2.5 levels correspond to lower air quality, while lower PM2.5 levels indicate better air quality.
- A graph can be generated for visualizing the predicted PM2.5 value in the given range using matplotlib and seaborn.
- An audio message can be created including the peak PM2.5 value,date,time and health precautions using gtts module and it can be broadcasted through Whatsapp using pyautogui.

High Level Overview

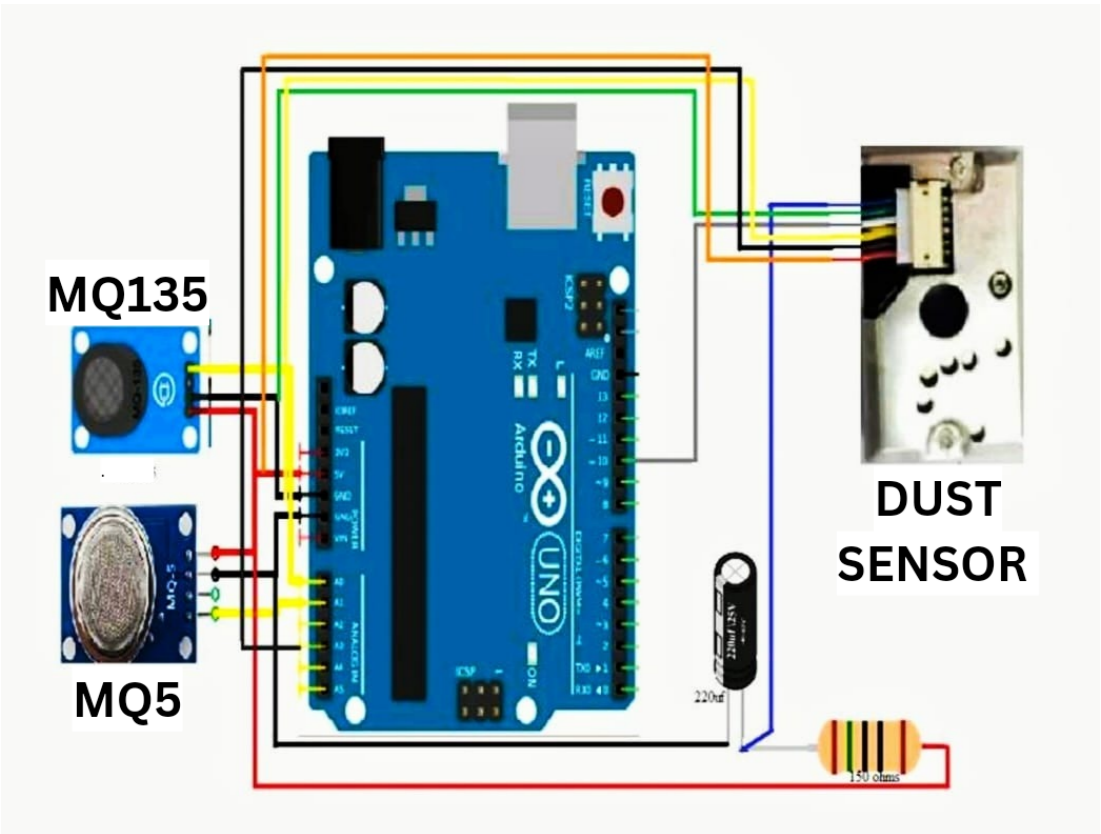

Figure: Hardware used for data collection

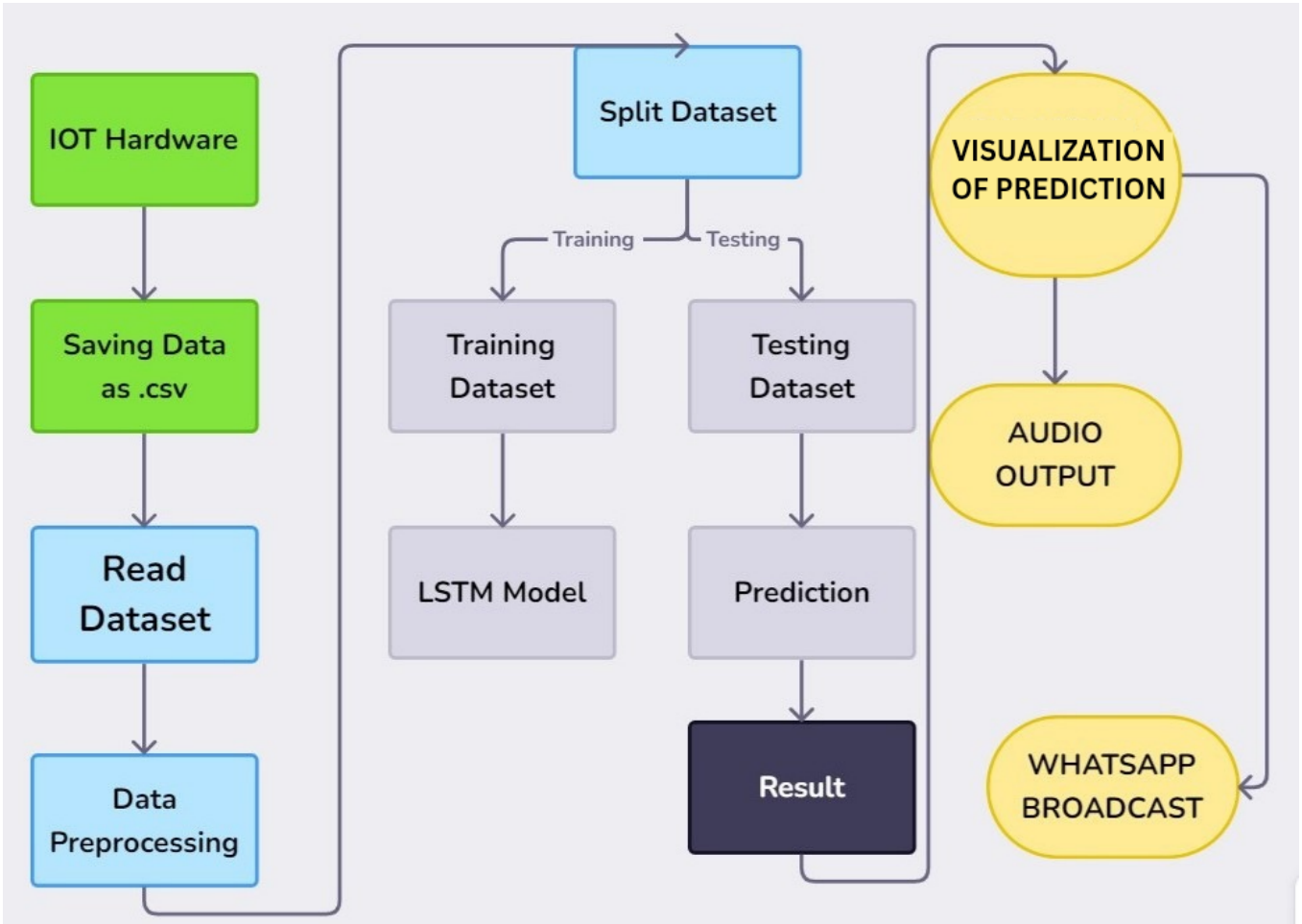

Figure: Block Diagram of Proposed System

Modules

- Collected Dataset:
  - Sensors deployed in locality measures air quality parameters like PM2.5, NO2, CO, O3, and other pollutants.
  - Timestamps accompany each measurement to monitor changes in air quality over time.
- Data Storage:
  - The collected data is stored in a CSV file, ensuring easy accessibility and compatibility for model training.
  - Data preprocessing involves tasks like cleaning, handling missing values, and outlier detection to ensure dataset quality and reliability.

- Model Training :
  - Train the chosen model using the collected dataset.
  - Feed input features and corresponding target values (air quality parameters) to help the model learn patterns and relationships.
- Model Testing and validation:
  - Tested the model using testing data and forecasted the air quality of upcoming days and plotted the predicted graphs of PM2.5.
  - Performance of model is validated using mean absolute error.

- User Interface Design:
  - A simple interface which takes range of dates as input and shows the graph of predicted PM2.5 value and a small dialogue box below the graph displaying the peak PM2.5 value and the date on which it is shown.
  - A voice output of the predicted peak PM2.5 value,date,time and health advisory will be given.
  - This audio file will be broadcasted through Whatsapp group (TRIVANDRUM PM2.5 ALERT!).

Tools and Methodologies Used

- Python version 3.6
- Jupyter notebook along with pandas,numpy.
- Visual Studio Code
- flask
- Single-step LSTM (Long-Short Term Memory)

- LSTM:
  - Gather historical air quality data, including PM2.5 measurements, temperature, humidity, and wind speed.
  - Normalize the data to ensure consistency across features and split it into training and testing sets.
  - Define the sequence length, determining the number of time steps to consider in the past for prediction.

- Construct the LSTM architecture with layers:
- Input layer to accept historical data sequences.
- LSTM layer(s) to learn temporal dependencies.
- Fully connected (Dense) layer to produce the output prediction.
- Compile the model by specifying the loss function (e.g., Mean Squared Error), optimizer (e.g., Adam), and optional evaluation metrics (e.g., Mean Absolute Error).

- Train the LSTM model on the training data.
- Adjust model weights using backpropagation to minimize the loss function.Iterate over the training data for multiple epochs until convergence.
- Make predictions using the trained model on validation sequence using Mean Absolute Error.
- Visualize predicted PM2.5 values against actual values to assess accuracy.

- keras(Tensorflow): facilitates the creation and deployment of precise PM2.5 predictive models by offering a user-friendly API for constructing, training, and evaluating LSTM networks.

- Matplotlib and Seaborn: Seaborn enhances Matplotlib by simplifying the creation of visually appealing statistical visualizations showcasing PM2.5 model predictions and trends.

- gtts(Google-Text-to-Speech):The gtts module enables the project to convert peak PM2.5 value and date information into speech, providing auditory feedback to enhance accessibility and usability.

- Pyautogui: Pyautogui automates interactions with the WhatsApp Desktop UI, facilitating the seamless sending of audio files containing health advisory to contacts or groups, thereby increasing efficiency and reducing manual involvement.

Result and Analysis

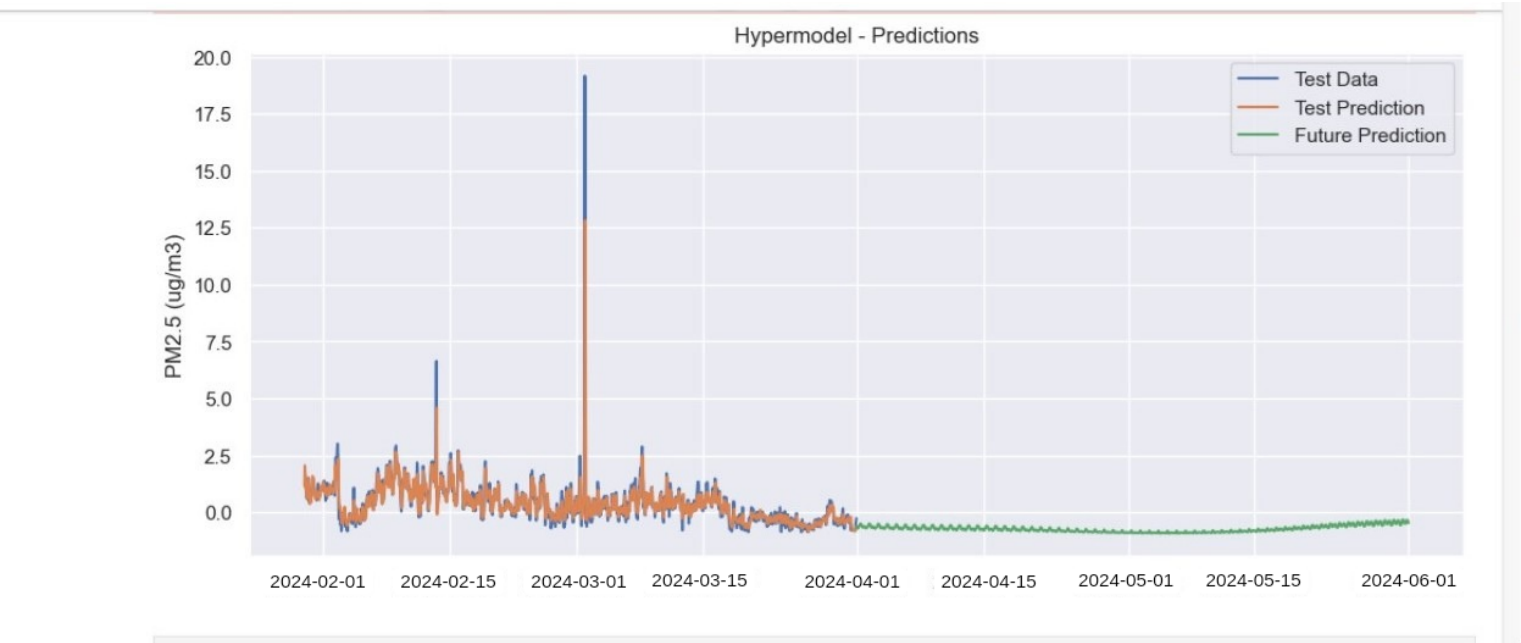

Figure: Graph showing relation b/w test prediction and future prediction

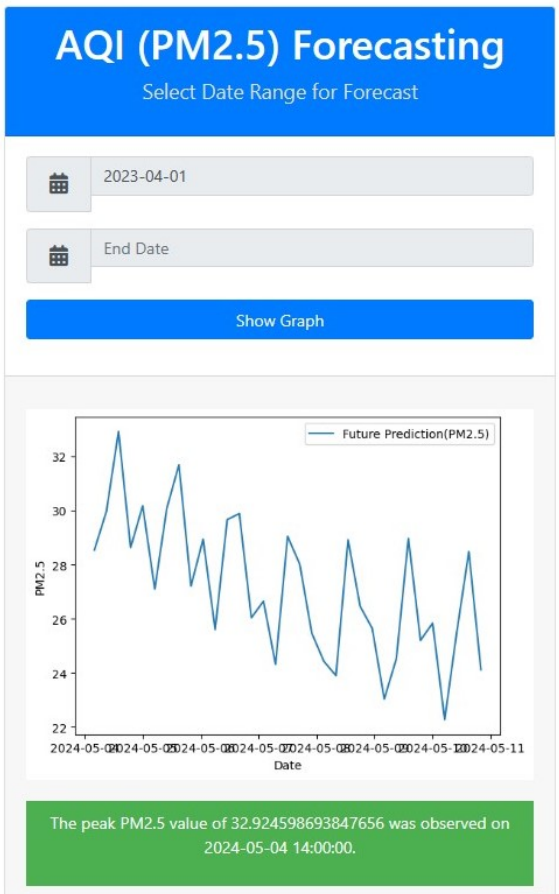

Figure: Graphical User Interface

- The graph shows how PM2.5 concentration varies over time, with dates plotted on the x-axis and PM2.5 levels on the y-axis. PM2.5 refers to tiny particles in the air that contribute significantly to air pollution.
- Generally, as PM2.5 concentration increases,, indicating poor air quality, and conversely, as PM2.5 levels decrease, signaling better air quality conditions.

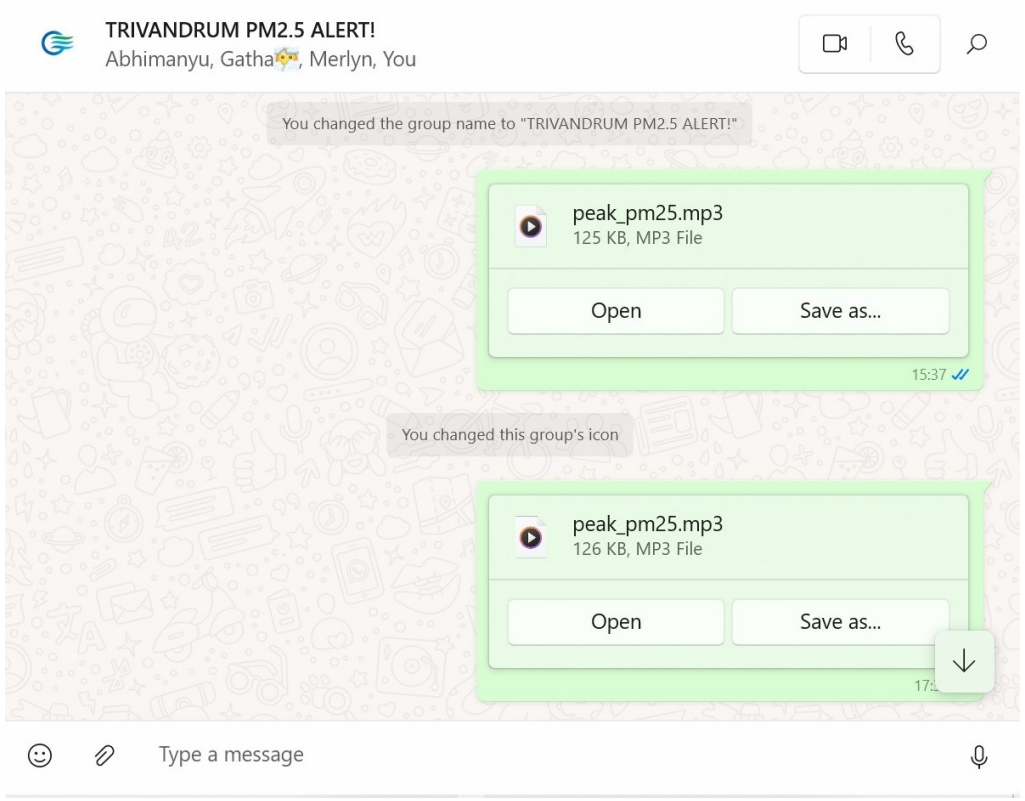

Figure: Whatsapp Alert!

- Utilize text-to-speech (TTS) technology to convert relevant informaton such as peak PM2.5 value,date and time at which it shown and health advisory into speech.
- Enhance user experience by providing auditory cues for accessing important data insights.
- Send voice notes to users via WhatsApp alerts, informing them about maximum PM2.5 values and accompanying health advisories.
- Utilize WhatsApp as a communication channel to deliver proactive notifications, empowering users to take necessary measures to safeguard their health.

Analysis(contd...)

Accuracy		Precision	
LSTM	SVR	LSTM	SVR
0.80	0.75	0.84	0.80

Figure: Comparison of results of LSTM and SVR models

- The prediction accuracy of the LSTM and SVR based model were calculated as 80% and 75%respectively.
- The best Precision 84% for LSTM and 80% for SVR.
- Therefore, this scores have shown that LSTM based model outputs has a better performance than the SVR model.

Conclusion

- Improved Predictions: Advanced technologies like LSTM networks enhance prediction accuracy and anticipate future air quality trends.
- Crucial Insights: This proactive approach provides healthcare providers, policymakers, and the public with important information for timely interventions.
- Reducing Health Impact: By acting on these insights, we can minimize the impact of air pollution on public health.

References

1. İbrahim KÖK, Mehmet Ulvi ŞİMŞEK, Suat ÖZDEMİR, *A deep learning model for air quality prediction in smart cities*, IEEE International Journal on Big Data, 2022
2. S.V.Vasantha, K.Ramesh, J.Vaishnavi, *Air Quality Prediction System using ML and DL Techniques*, IEEE North Karnataka Subsection Flagship International Conference, 2020
3. Deepu B.P, Dr. Ravindra Rajput, *Air Pollution Prediction Using Machine Learning*, IRJET volume 10, 2022
4. S. Venkateswarlu1, D. B. K. Kamesh1, J. K. R. Sastry2 and Radhika Rani, *Text to Speech Conversion*, Indian Journal of Science and Technology, October 2016.
5. Dr. V. Vasanthi, Vidhyabaran Rajkanna . T, *Python Based Gesture Control Using Pyautogui*, JETIR, April 2023.

Thank you.