AIR QUALITY PREDICTION USING DEEP LEARNING MODELS

A MINOR PROJECT REPORT

Submitted by

SASHANK DONAVALLI [RA2011026010234]
HARSHITHA KAMBHAM [RA2011026010235]

Under the guidance of

Dr. M. Dhilsath Fathima
ASSISTANT PROFESSOR, DEPARTMENT OF CINTEL

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

of

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

S.R.M.Nagar, Kattankulathur, Chengalpattu District

NOVEMBER 2023

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Under Section3 of UGC Act,1956)

BONAFIDE CERTIFICATE

Certified that 18CSP107L minor project report titled "**AIR QUALITY PREDICTION USING DEEP LEARNING MODELS**" is the bonafide work of "SASHANK DONAVALLI (RA2011026010234), HARSHITHA KAMBHAM (RA201026010234)" who carried out the minor project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of whicha degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr. M. Dhilsath Fathima
GUIDE
ASSISTANT PROFESSOR
Department of Computational Intelligence

SIGNATURE

Dr.R. Annie Uthra
HEAD OF THE DEPARTMENT
PROFESSOR
Department of Computational Intelligence

SIGNATURE

Dr. C. Amuthadevi
PANEL HEAD
ASSOCIATE PROFESSOR
Department of Computational Intelligence

ACKNOWLEDGEMENT

We express our humble gratitude to **Dr. C. Muthamizhchelvan**, Vice-Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support. We extend our sincere thanks to Dean-CET, SRM Institute of Science and Technology, **Dr. T. V. Gopal**, for his invaluable support. We wish to thank **Dr. Revathi Venkataraman**, Professor and Chairperson, School of Computing, SRM Institute of Science and Technology, for her support throughout the project work.

We are incredibly grateful to our Head of the Department, **Dr. R. Annie Uthra**, Professor, Department of Computational Intelligence, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

We want to convey our thanks to our Panel Head, **Dr. C. Amuthadevi**, Associate Professor and Panel Members, Department of Computational Intelligence, SRM Institute of Science and Technology, for their inputs during the project reviews and support.

We register our immeasurable thanks to our Faculty Advisor, **Dr. Velliagiri S**, Assistant Professor, Department of Computational Intelligence, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to our guide**, Dr. Dhilsath Fathima M**, Assistant Professor, Department of Computational Intelligence, SRM Institute of Science and Technology, for providing us with an opportunity to pursue our project under his mentorship. She provided us with the freedom and support to explore the research topics of our interest. Her passion for solving problems and making a difference in the world has always been inspiring. Finally, we would like to thank our parents, family members, and friends for their unconditional love, constant support and encouragement.

SASHANK DONAVALLI [Reg No:RA2011026010234]

HARSHITHA KAMBHAM [Reg No:RA2011026010235]

ABSTRACT

Air Pollution is a significant concern in today's world for both the public and the environment. Monitoring and predicting air pollutant concentrations are essential for effective pollution control and management. This study focuses on the development of a time series prediction model for multiple pollutants in the air using advanced deep-learning techniques. The time series data is taken into account, capturing temporal dependencies and seasonal patterns that are inherent in air quality datasets.

This paper ties to explores the use of various deep learning architectures, including Recurrent Neural Networks (RNNs) and other hybrid models. This research highlights the effectiveness of deep learning models in air quality prediction and their potential to provide accurate and timely information for air quality monitoring and management. These models not only enhance air quality prediction but also offer the potential for real-time monitoring and early warning systems for air quality management. Furthermore, the study explores feature engineering techniques, data pre-processing, and hyperparameter tuning to optimize the deep learning models. Cross-validation and evaluation metrics such as MAE, R2 and RMSE are employed to assess the model's performance.

In conclusion, this project shows how deep learning models can effectively increase the precision of air quality prediction, offering useful information to the public, environmental organizations, and policymakers to help them make decisions about air quality and how it affect environment and humans as a whole. The results of this study have the potential to considerably advance monitoring and control efforts for air quality.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

RNN : Recurrent Neural Network

BiLSTM : Bidirectional Long Short Term Memory

GRU : Gated Recurrent Unit

LSTM : Long Short Term Memory

AQI : Air Quality Index

DL : Deep Learning

EWS : Early Warning System

FED : Feature Engineering And Designing

CHAPTER 1

INTRODUCTION

The growing problem of urban air pollution significantly hampers the City's capacity to flourish sustainably and establish an ecological civilization. The lives, productivity, and health of the people are all impacted by the quality of the air. Accurate and fast air quality evaluation is necessary to solve this challenge. By taking into account a variety of pollutants, such as (PM2.5 and PM10),(O_3) ,(CO), (SO_2), and nitrogen dioxide (NO_2), the quality of air is predicted using various deep learning models. However, precise prediction of air quality remains a complex task due to the intricate interplay of environmental factors and pollutants.

Traditional methods for air quality prediction often rely on statistical and machine learning models, which may struggle to capture the intricate patterns and dependencies in air quality data. In recent years, the advancement of deep learning models has shown remarkable potential for improving the accuracy and reliability of air quality predictions. The goal of this work is to investigate the use of deep learning models in air quality prediction. Our objective is to use these model's ability to more accurately predict, interpret, and maintain a focus on air quality situations. Deep learning's primary benefit is its capacity to identify geographical and temporal relationships in air quality data, which offers a more thorough and precise understanding of air quality dynamics. We may be able to overcome the drawbacks of traditional air quality prediction methodologies by utilizing deep learning techniques. This could lead to the development of more potent early warning systems, decision support tools, and enhanced air quality management.

Deep learning models used in detecting air quality index in this project are RNN,LSTM,GRU, and BiLSTM. Thechoice of deep learning model depends on the specific characteristics of the air quality data, the size of the dataset, and the required prediction horizon. It's important to pre-process the data properly, split it into training, validation, and test sets, and fine-tune the model's hyperparameters to achieve the best performance.

1.1 Existing System

In the past, researchers have experimented with a variety of ML models (including logistic regression, naive bias, RF and support vector machines). Using machine learning techniques like K-Means Clustering, these models also employ classifications such as bad, satisfactory, good, and very good for air quality. Convolutional and recurrent neural networks (RNNs), two examples of deep learning models, have become more and more well-liked due to their capacity to identify intricate temporal patterns in data on air quality. Numerous studies have shown how crucial it is to take into account both the spatial and temporal dimensions of air quality. Temporal dependencies have been captured through the use of time series analysis techniques such as autoregressive model and LSTM networks. Researchers have examined the significance of diverse input features in the prediction of AQI. It has been discovered that weather-related factors, like wind speed and temperature, are important for forecasting air quality. Geographical aspects have also been taken into account, such as closeness to sources of pollution.

Previous studies have faced difficulties with missing data handling, overfitting of the model, and the requirement for environmental science domain expertise. Concerns have also been raised concerning non- stationary data and the interpretability of complicated models. Some key challenges faced in the past related to air quality index include data quality and availability, feature engineering, temporal and spatial variability, overfitting and generalization, real time prediction, data integration, handling extreme events, updating according to environmental changes and climate variability and regulatory considerations. Addressing these challenges often requires interdisciplinary collaboration between data scientists, environmental scientists, meteorologists, and policymakers. Developing robust and accurate AQI prediction models is crucial for safeguarding public health and the environment, but it requires a careful consideration of these challenges throughout the research process.

1.2 Problem Statement

Medical Concern over air pollution is growing since it has a big effect on the environment and public health. For the general public, local government agencies, and businesses to be aware of the air quality conditions and be able to take the necessary precautions to reduce the risks to their health and the environment, it is essential to predict and monitor the Air Quality. The objective of this project is to use historical air quality and meteorological data to build a deep learning-based system for accurate and real-time air quality prediction. High-resolution air quality prediction forecasts will be provided by the system, enabling people and decision- makers to make well-informed decisions that will save pollution and safeguard their health.

Some key objectives need to be achieved in this project are:

- **Model Comparison:** Compare the performance of various deep learning models, such as Recurrent Neural Networks (RNNs), Long Short Term Memory networks (LSTMs), and hybrid models, in predicting AQI. Determine which model architecture yields the most accurate predictions.

- **Data Pre-processing:** Develop and implement effective data pre-processing techniques to handle missing values, outliers, and noise in air quality and meteorological data. Assess how different pre-processing methods impact model performance.

- **Feature Engineering:** Investigate the importance of different air quality parameters (e.g., PM) and meteorological variables (e.g. wind speed, temperature, wind direction) in AQI prediction. Explore feature selection and extraction methods to identify the most relevant features.

- **Hyperparameter Tuning:** Optimize hyperparameters for each deep learning model to enhance prediction accuracy. Evaluate the sensitivity of model performance to hyperparameter changes.

- **Temporal Analysis**: Investigate the temporal patterns and trends in air quality data, including daily, weekly, and seasonal variations. Develop models capable of capturing and leveraging temporal dependencies in air quality data.

- **Spatial Analysis:** Explore the spatial distribution of air quality data by considering data from multiple monitoring stations in different geographic locations. Develop models that can account for spatial correlations and variabilities in air quality.

- **Performance Metrics:** Evaluate model performance using appropriate metrics, including MAE, RMSE, R2, and others. Compare the models based on these metrics.

- **Scalability and Generalization:** Assess the scalability of the selected models and their ability to generalize across different regions and timeframes. Investigate any limitations or biases that may arise in model generalization.

The successful implementation of this deep learning-based air quality prediction system will empower residents, local authorities, and industries to make informed decisions, reduce air pollution, and protect public health. It will contribute to the mitigation of environmental and health risks associated with air pollution.

1.3 Technical Requirements

These requirements encompass various aspects, including hardware, software, data, and infrastructure. Collaborating with domain experts and stakeholders is essential to ensure that the system meets the specific needs of the region and adheres to environmental standards and regulations.

Hardware Requirements-

- Laptop / PC with any OS (Window 7 or later, Mac OS (any version), Linux (any version)
- RAM storage of minimum 8GB and Hard Disk storage of minimum 128GB is required.
- Constant internet connection
- Laptops or PCs with an i3 processor or better.
- Virus-free environment

Software Requirements-

- Operating System: Windows 7/8/10
- Libraries: numpy, tensorflow, sklearn,pandas.
- Programming Language: Python
- IDE/Workbench:CoLab

CHAPTER 2
LITERATURE SURVEY

In research paper [1] suggests an environmental quality prediction model based on LSTM, in light of the state's increased focus on environmental governance in recent years and the ongoing decline in air quality. With the use of temperature, PM2.5, PM10, SO2, wind direction, NO2, CO, and O3, this study forecasts the Air Quality Index (AQI). The Environmental Protection Agency provides the statistics. First, this study presents the history, technological aspects, current status, and difficulties related to air environment monitoring. The environmental prediction model will then be introduced. Using LSTM, finally producing an AQI prediction and examine the prediction error. According to the findings, LSTM is a good predictor of the air quality index.

In research paper [2] to anticipate the hourly AQI, a number of approaches are used, including a linear model and cutting-edge methods like BPNN, CNN, GRU, LSTM, and BiLSTM. Experiments evaluating the performance of different strategies reveal that the BiLSTM provides the greatest results.

In paper [3] suggests a hybrid approach to indoor air pollution prediction using deep neural networks and fuzzylogic. We also train and forecast the dataset obtained by indoor sensors in Shanghai between November 2016 and March 2017 using PM2.5 pollution as an example. Using LSTM ,CNN-LSTM, and our suggested FL- CNN-LSTM network built on the PyTorch framework; created comparison tests to test the prediction process. According to the findings, deep neural networks that have been infused with fuzzy logic can offer improved predictability and interpretability for the intended use.

A Bayesian Optimization with Stacked Deep Learning-based Air Quality Index Prediction (BOSDL-AQIP) method was used in paper [4]. The effective identification and categorization of Air Quality (AQ) into many class labels is the aim of the BOSDL-AQIP technique. The proposed BOSDL-AQIP approach uses min-max normalization for data scaling in order to achieve this. Next, for the prediction phase, the BOSDL-AQIP system makes use of the Stacked Bidirectional Long Short-Term Memory (SBiLSTM) approach.

This research [5] proposes a hybrid air quality prediction model that combines deep neural networks and K- Means clustering. BiLSTM and fully connected neural networks make up the deep neural network with regressive computation capability. First, the research goal is the historical meteorological monitoring data of Qingdao City. Using the k-Means clustering technique, the meteorological data is separated into four categories based on the quarter.

The dataset utilized in the paper [6] is made up of hourly data on multiple air pollution types collected at multiple locations in multiple Indian cities. Decision Tree, SVM and RF, and were the models we used. A method for random forest categorization produces the most accurate results with a maximum accuracy of 74%, it outperformed the other methods. The present research will benefit from these findings, which will also direct future investigations.

In order to make the future air quality profile easier for the general public to understand, a new comprehensive evaluation approach called LSTM-Fuzzy is developed in paper [7], which combines the predicted value of LSTM with the fuzzy algorithm. The experiment results show that while the accuracy of UF is often stable in comparison to the DF, the accuracy of DF typically declines as the forecast time grows.

The paper [8] is to examine the quality of air in India, as well as the impact of seasonal variations and COVID-19 on air pollution levels and, consequently, the AQI. Full-scale analysis is carried out, accounting for various granularities, including daily, weekly, and monthly data. To achieve the best findings, this study extensively preprocesses the time series data for air quality. The findings showed that particulate matter, specifically PM 2.5 and PM 10, had the most influence on air quality. Analysis of the effects of the COVID-19-related countrywide shutdown as well as the seasonal variations on overall air quality have been done.

To achieve the AQI's short-term prediction in paper [9], a distributed neural network model with the AQI change rule included in the distributed neural network structure is used. Based on self-learning features, experimental results demonstrate that air quality prediction models based on big data and neural networks can indicate the development trend of air quality. It can offer a scientific foundation for the amount of urban air pollution and assist people in taking the proper actions for varying AQI levels. It also has a greater forecast accuracy.

In paper [10], The dataset was analyzed using the supervised machine learning technique (SMLT) to extract various information such as variable recognition, uni-variate, bi-variate, and multi-variate analysis, missing value treatment and analysis, data preparation and cleaning, and data representation. Additionally, to contrast and talk about several machine learning algorithms to find the best accurate algorithm that can execute a GUI-based user interface for predicting air quality.

In this paper [11], historical datasets were examined that included measurements of many air pollutants, including SO2, NO2, SPM, RSPM, and PM 2.5, for the years 1998 through 2020. These values of air pollutants in the dataset are then used to generate the air quality indices (AQI) to find the AQIs for upcoming years in various Indian cities. Future air quality predictions and model training are done using supervised machine learning techniques including Random Forest, Decision Tree, Logistic Regression, and Linear Regression.

This research paper [12] suggests a way to forecast pollution levels for any given day by integrating the Internet of Things (IoT). The paper demonstrates the use of sensor-derived data in smart air pollution monitoring systems to intelligent pollution control. A dataset is created through the ETL process in Jupyter notebook by gathering data from the air pollution sensor, which sends the data to the server via NET 6 REST API endpoint. The data is then placed in a SQL Server database along with additional weather data that is collected from REST API for that time of day. The prediction methods that will be employed are linear regression techniques.

In this paper [13], researchers built and put into action a portable, affordable, and tiny hardware platform for monitoring air quality that can be accessed remotely through a web application. Additionally, they have used a dataset and certain forecasting algorithms to predict each parameter over a given time frame.

This paper [14] combines meteorological data and PM 2.5 concentrations from nearby stations to forecast PM 2.5 hourly scale concentrations using machine learning models, namely Decision Tree (DT), Random Forest (RF), and Gradient Boosting Regressor (GBR). The study region, Beijing, China, is where the dataset was gathered. With an R2 value ranging between 0.9 and 0.97, the trials have demonstrated that the gradient boosting regressor model delivers superior predictive precision than the other models provided for forecasting hourly PM 2.5 concentrations. A viable and reasonably priced method for reliably predicting PM 2.5 concentrations is provided by this work.

In paper [15] , well-designed spectral graph convolution with a second-order Chebyshev polynomial is created after a spatial adjacency network structure based on correlations of air monitoring stations is built. The affects from distant stations can be captured using this spatial module. LSTM attention mechanism is used to the input of spatial characteristics from spatial module and node features to capture temporal dependence. The hourly scaled dataset from 12 air quality monitoring sites in Shanghai, which includes spatiotemporal characteristics, air pollutant concentrations, and meteorological parameters, is used to apply this model.

CHAPTER 3

SYSTEM ARCHITECTURE AND DESIGN

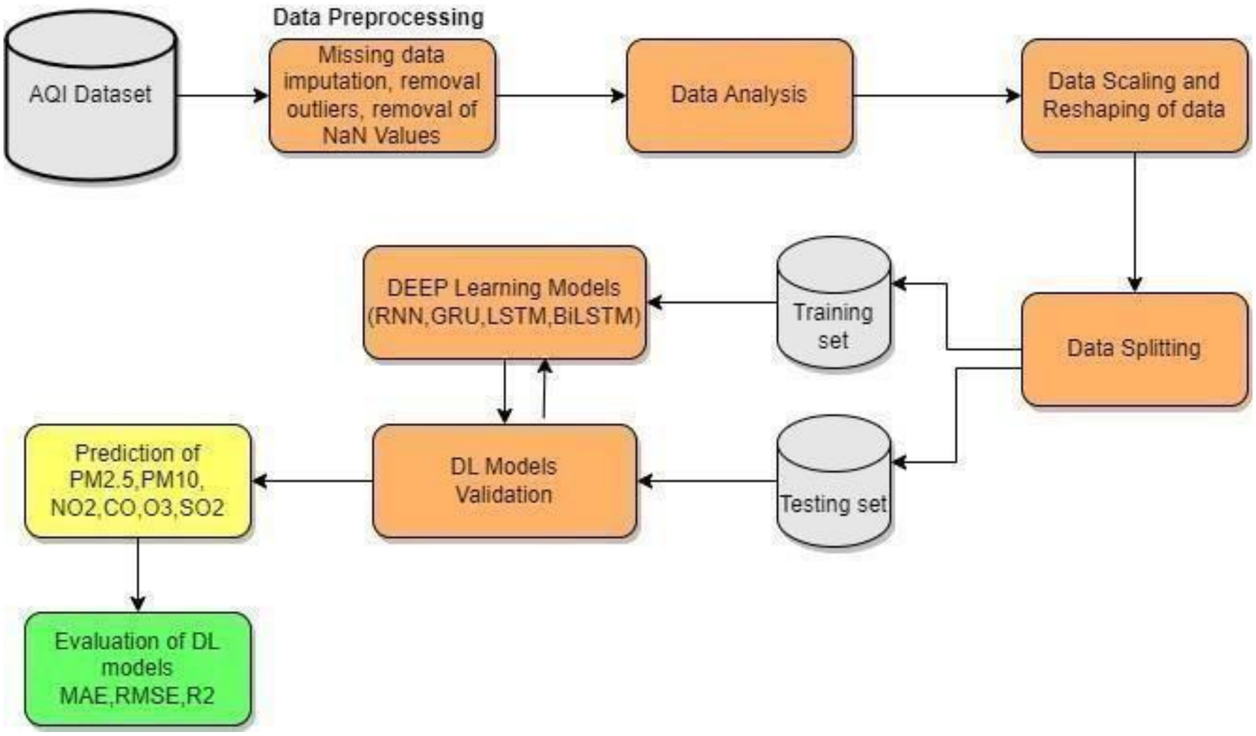

Fig 3.1 System Architecture

The system architecture(Fig3.1) for an air quality monitoring and prediction system typically involves several components and data sources that work together to provide real-time or forecasted air quality information.

Components of the architecture include:

1. **Data Sources:**

- Air Quality Monitoring Stations: These are ground-based stations equipped with sensors to measure pollutant concentrations, such as PM2.5, O3, PM10, NO2, SO2 and CO.
- Meteorological Data: Understanding weather conditions is crucial for determining air quality. These conditions include wind direction, humidity, wind speed, temperature, and atmospheric pressure.
- Historical Data: A historical database of pollutant levels, weather conditions, and AQI values is used for training and validation.

2. Data Collection: Gather historical air quality data, meteorological data, and geographical information from monitoring stations from Kaggle. Pollutants with respect to station.

3. Data Preprocessing: In Data preprocessing, initial steps are taken to clean the data. All the missing values in the given dataset are replaced by the mean values. The outliers are removed to ensure data integrity and quality are maintained.

4. Data Correlation: The direct cause of the drop in air quality is pollution. Air circulation also helps to disperse toxins in the environment. There is a high correlation (Fig 3.2) among the pollutants. The six pollutants namely O3, CO, NO2, SO2,PM10, and PM2.5 are used for prediction.

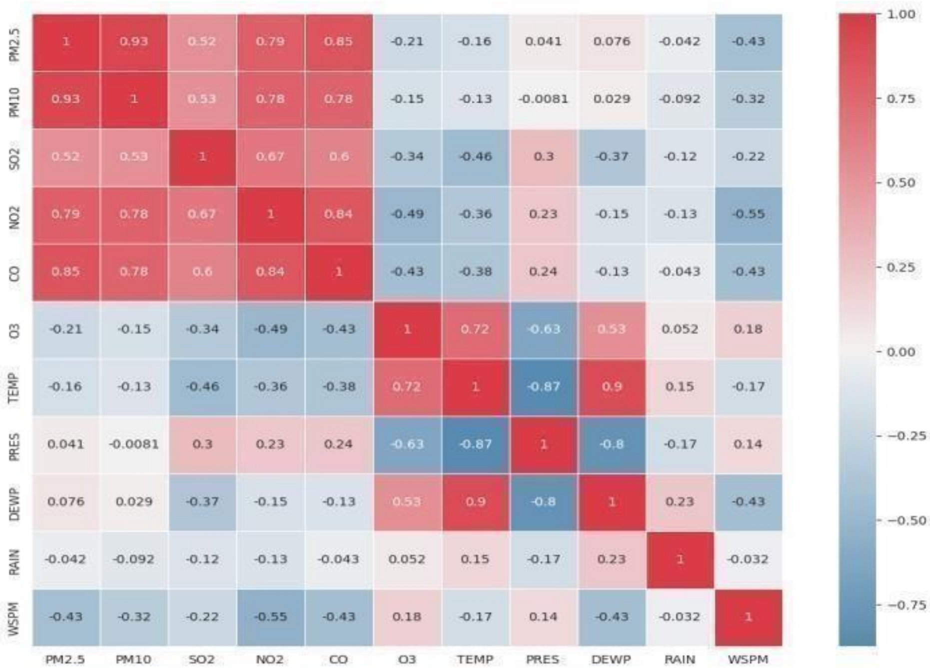

Fig 3.2 Correlation Analysis

5. Data Splitting: Split the data into testing, training and validation data for model development and evaluation.

6. Model Selection: Deep learning models such as RNN, LSTM, GRU, Bi-LSTM are used for predicting the 6 pollutants.

7. Model Training: Train the selected models on the training data using appropriate loss functions and Adam optimizer is used.

8. Model Evaluation: The model's performance using metrics such as MAE, RMSE,R2.

9. Visualization: Visualize model predictions and compare each model prediction of 6 pollutants to actual values to gain insights and identify any discrepancies.

10. Monitoring and Maintenance: Constantly assess the model's performance and make updates when new information becomes available or the environment shifts.

Each of these steps is crucial for developing an effective air quality prediction system that can provide valuable information for public health, environmental management, and urban planning.

3.1 Use case Diagram:.

A use case diagram Fig 3.1.1 visually represents user interactions with a system, defining functionalities and aiding communication. It guides requirements analysis, system design, and project planning in software development.

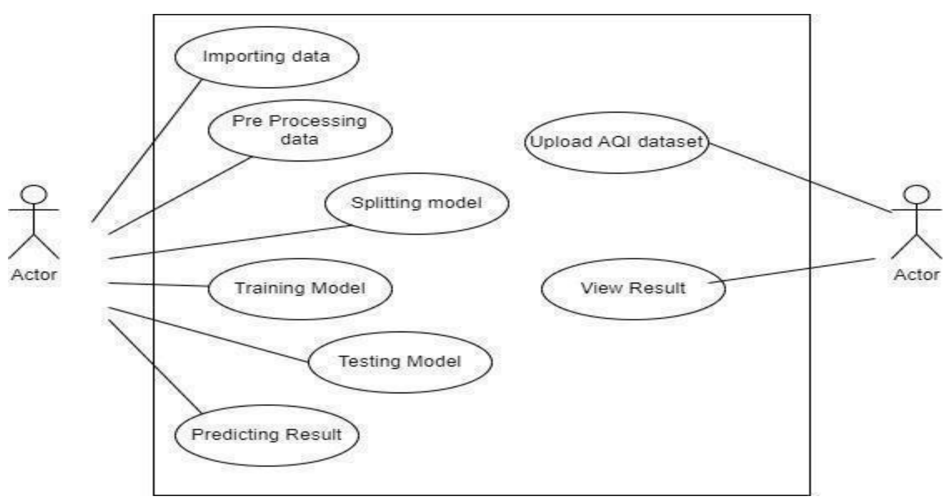

Fig 3.1.1 Use case diagram

3.2 Sequence Diagram

Sequence diagram in Fig 3.2 visually represent the chronological flow of interactions between objects or components in a system, aiding in understanding, designing, and documenting the dynamic aspects of the model’s processes.

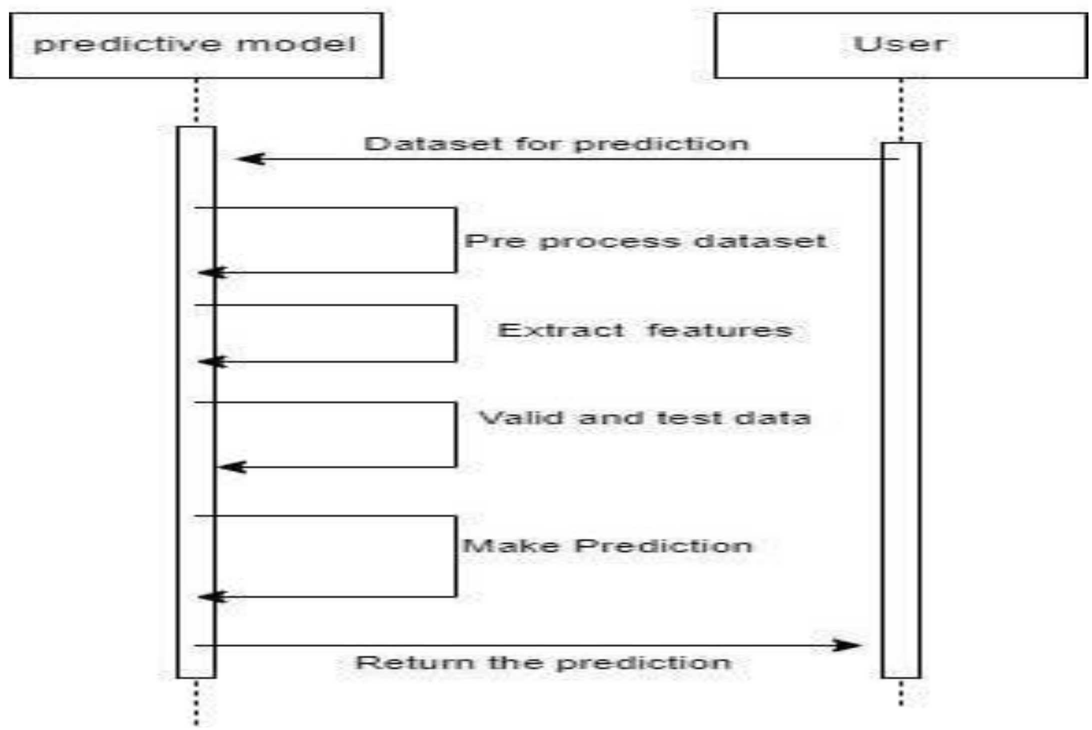

Fig 3.2.1: Sequence diagram

3.3 Activity Diagram

Activity diagram Fig 3.3.1 visually represent the flow of activities in a system, aiding in understanding and modeling complex processes. It illustrate actions, decisions, and parallelism, enhancing system design and communication.

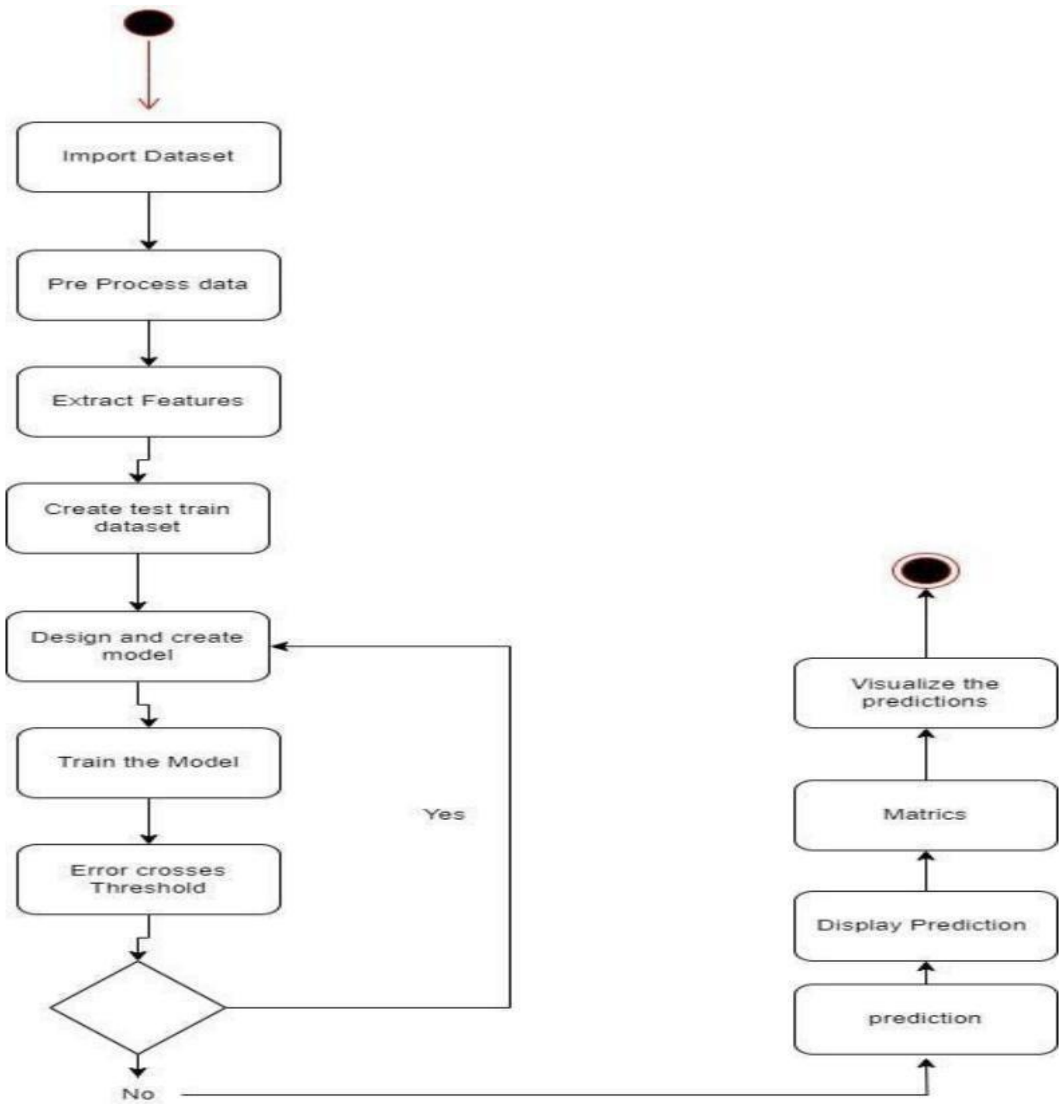

Fig 3.3.1 Activity diagram

3.4 Component Diagram

A component diagram in Fig 3.4.1 visually represents the high-level structure and relationships between system components. It aids in system design, illustrating the modular organization and dependencies for development teams.

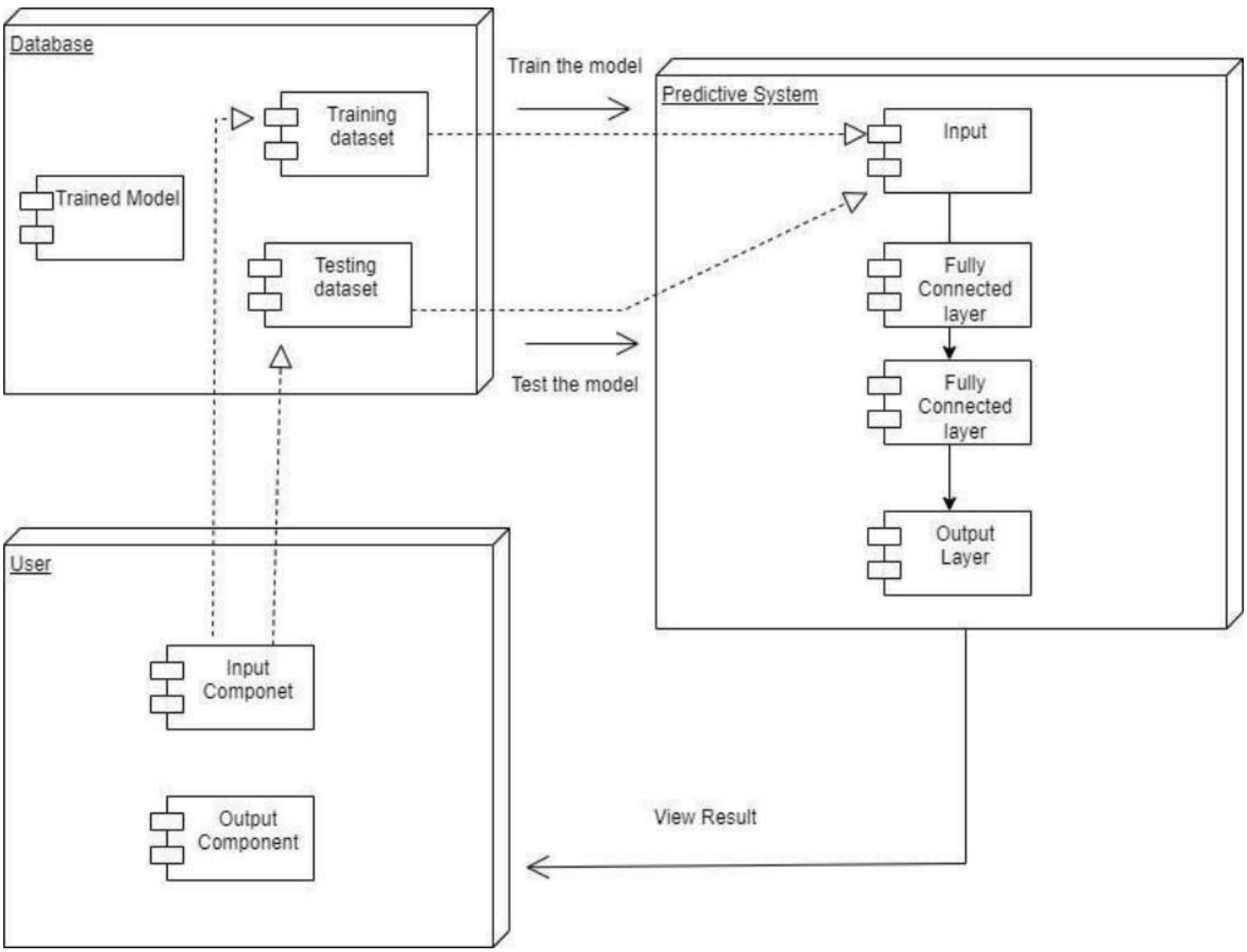

Fig 3.4.1 Component diagram

CHAPTER 4

METHODOLOGY

The Conventional techniques for air quality prediction frequently depend on machine learning and statistical models, which may find it difficult to identify the complex relationships and patterns in air quality data. The development of deep learning models in recent years has demonstrated remarkable potential for enhancing the precision and dependability of air quality predictions. Examining the application of deep learning models to air quality prediction is the aim of this work. Our goal is to make use of these model's capacity for more precise interpretation, prediction, and attention to air quality conditions. The main advantage of deep learning is that it can be used to find spatial and temporal patterns in air quality data, providing a more accurate and comprehensive understanding of air quality dynamics. This could lead to the development of more potent early warning systems, decision support tools, and enhanced air quality management.

This study proposes to use four deep learning models for air quality prediction:

3.4.1 RNN
3.4.2 LSTM
3.4.3 GRU
3.4.4 BiLSTM

Using these prediction models the paper proposes to predict all six pollutants namely PM2.5,CO,NO2,PM10,SO2,O3. Pollutant prediction is crucial for a number of reasons, including the fact that it has a big impact on the environment, public health, and many different facets of society.

4.1 Recurrent Neural Networks (RNN's)

By identifying the temporal relationships in the air quality data, recurrent neural networks(RNNs) are utilized to forecast the air quality index (AQI). RNNs use historical data to predict future AQI levels, assisting in environmental preservation and public health planning. They are

useful tools for AQI forecasting because they are excellent at handling time-series data with intricate temporal patterns and varying sequence lengths. In the below RNN architecture (Fig 4.1) an input layer that accepts input data with the shape of (None, 1, 33), is where the model starts. This shape implies that the model has 33 characteristics and one time step for input sequences.
Variable batch sizes are often represented by the "None" dimension. Following the input layer comes a Simple RNN (Simple Recurrent Neural Network) layer with 128 units. Recurrent neural network layers called Simple RNN are utilized for sequence data. Weights and biases are among the 20,736 parameters in the Simple RNN layer. A dense layer with 128 units and the label "dense_2" follows the Simple RNN layer. This layer contains 16,512 parameters and is fully linked. Another dense layer with the name "dense_3" and six units is present.

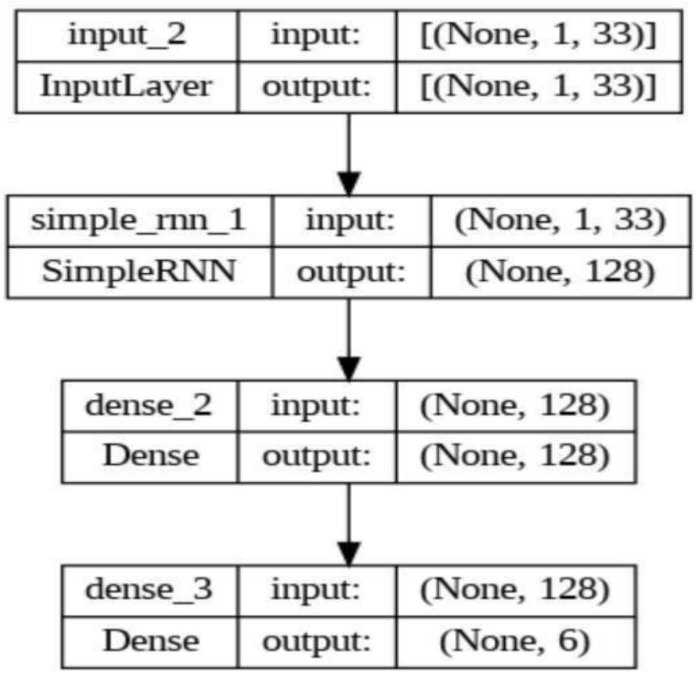

Fig 4.1 RNN Model

4.2 Long Short-Term Memory (LSTM)

LSTM architecture is applied to predict Air Quality Index (AQI). LSTMs excel in capturing complex temporal patterns in air quality data, making them valuable for forecasting pollution levels. These recurrent neural networks maintain long-term dependencies, enabling accurate predictions of AQI, aiding environmental and health management. The model’s Input layer that accepts input data with the shape of (None, 1, 33). (Fig 4.2) implies that the model anticipates 33 characteristics and one time step for input sequences. Variable batch sizes are often represented by the "None" dimension. A dense layer of long short-term memory (LSTM) with 128 units follows the input layer. Weights and biases are among the 82,944 parameters in the LSTM layer. Long-term dependencies in sequences can be captured by the LSTM layer, a form of recurrent neural network layer that is employed for sequence data. A dense layer with 128 units called "dense_8" follows the dense layer. With 16,512 parameters, this layer is dense and fully connected. Another dense layer with the designation "dense_9" and six units is present at the end, and it is most likely the output layer. 774 parameters are in this layer.

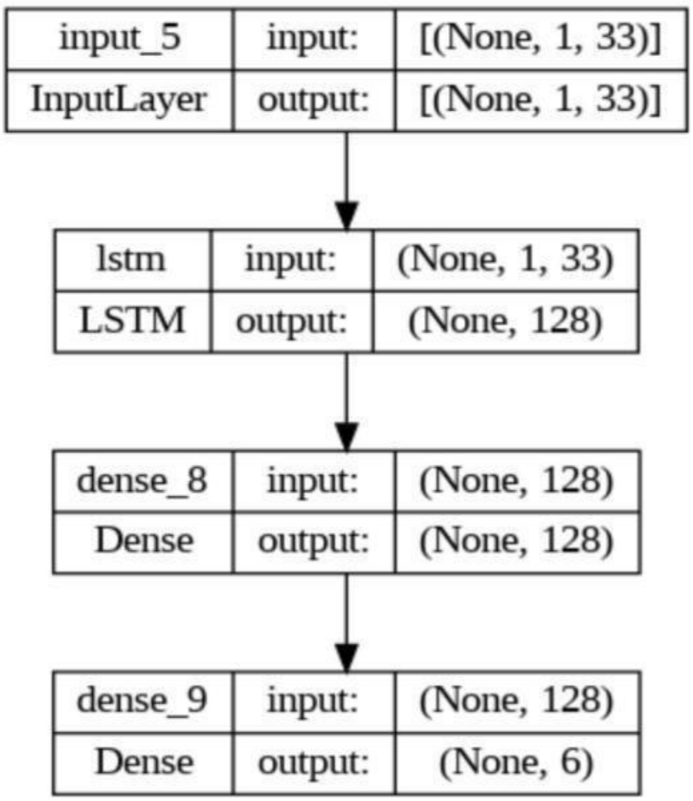

Fig 4.2 LSTM Model

4.3 Gated Recurrent Unit (GRU)

A GRU is similar to that of a RNN architecture used in deep learning for sequential data processing. It is a variant of the traditional RNN that addresses some of the issues related to vanishing gradients and long- term dependencies. The model's first layer is an Input layer that accepts input data with the shape of (None, 1, 33). This shape implies that the model anticipates 33 characteristics and one time step for input sequences. Variable batch sizes are often represented by the "None" dimension. In (Fig 4.3), a dense layer with 128 units and follows the input layer. Weights and biases are among the 62,592 parameters in the GRU layer. Recurrent neural network layers called GRUs are employed for sequence data. A dense layer with 128 units and the name "dense_6" follows the GRU layer. With 16,512 parameters, this layer is dense and fully connected. Another dense layer with the name "dense_7" and six units is present, and it is most likely the output layer. Another dense layer with the name "dense_7" and six units is present, and it is most likely the output layer.

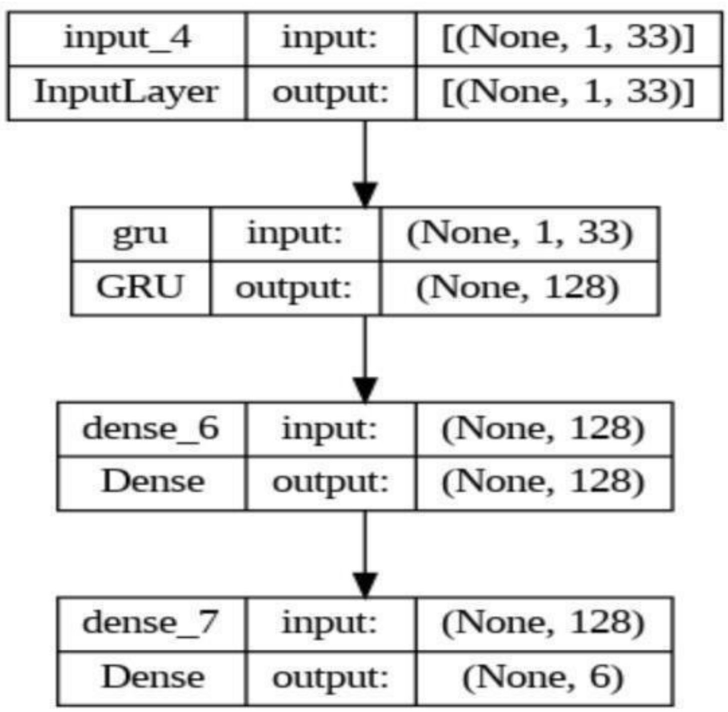

Fig 4.3 GRU Model

4.4 Bidirectional LSTM

Bi-LSTM (Bidirectional Long Short-Term Memory) deep learning model is used to forecast the air quality index (AQI). To effectively anticipate AQI, which is influenced by complex temporal dependencies and climatic factors, it takes advantage of its bidirectional nature to collect both future and past context. In Fig 4.4 input layer accepts input data with the shape of (None, 1, 33), is where the model starts. This shape implies that the model anticipates 33 characteristics and one time step for input sequences. Variable batch sizes are often represented by the "None" dimension. A Bidirectional layer with the label "bidirectional" follows the input layer. Sequence data is frequently processed in both forward and backward directions using bidirectional layers. There are 165,888 parameters in this bidirectional layer. An additional layer called "multi head attention" with 131,712 parameters follows the bidirectional layer. For attention-based operations and sequence processing, models like Transformers use a mechanism called multi- head attention. After the Flatten layer, there is a dense layer with 128 units . 32,896 parameters make up this dense layer. Finally, there is another dense layer with and six units; this layer is the output layer.

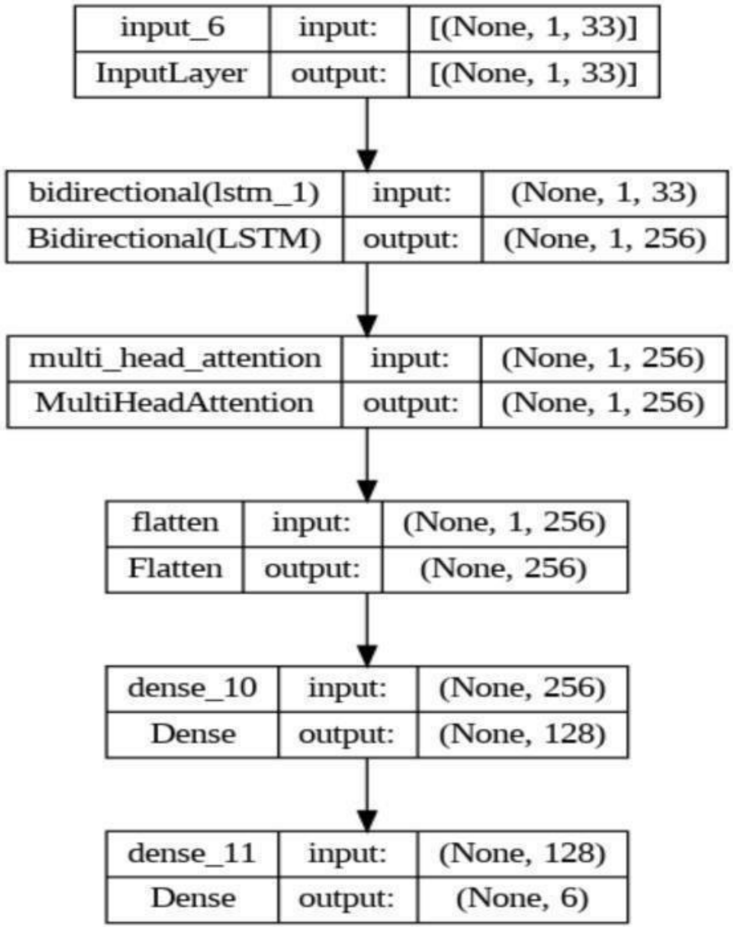

Fig 4.4 BiLSTM Model

CHAPTER 5

CODING AND TESTING

```
epochs = 100
batch_size = 32
learning_rate = 0.001 timesteps = train_X.shape[1] input_dim = train_X.shape[2]
output_dim=6
```

RNN model:

```
inputs = Input(shape=(timesteps, input_dim))
rnn_out = SimpleRNN(128,activation='relu',return_sequences =

False)(inputs) dense = Dense(units=128,

activation='relu')(rnn_out)

output = Dense(units=output_dim,

activation='relu')(dense) rnn =

Model(inputs=inputs, outputs=output)

rnn.compile(loss='mse',
    optimizer=keras.optimizers.Adam(learning_rate=learning_rate),
    metrics=['accuracy'])

# plot_model(model, to_file="Bi-LSTM-
MultiheadAtt.png", #         dpi=300,
show_shapes=True)

path_checkpoint =
"model_checkpoint_MultiAttLSTM_aqiPre.h5"
es_callback = keras.callbacks.EarlyStopping(
    monitor="val_loss",
    min_delta=0, patience=10
)

modelckpt_callback =
    keras.callbacks.ModelCheckpoint(
    monitor="val_loss",
    filepath=path_c
```

```
    heckpoint,
    verbose=0,
    save_weights_o
    nly=True,
    save_best_only
    =True,
)

# fit network
history = rnn.fit(train_X, train_y, epochs=epochs, batch_size=batch_size,
validation_data=(valid_X, valid_y), #        callbacks=[es_callback, modelckpt_callback]
                    )
visualize_loss(history, "Training and Validation Loss")
visualize_acc(history, "Training and Validation Accuracy")
```

In Fig 5.1 Training and validation loss graph for RNN model is plotted. These graphs help in the model improvement by showing convergence and detecting overfitting, ensuring optimal performance and generalization. Fig 5.2 the graph for validation and accuracy training for RNN.

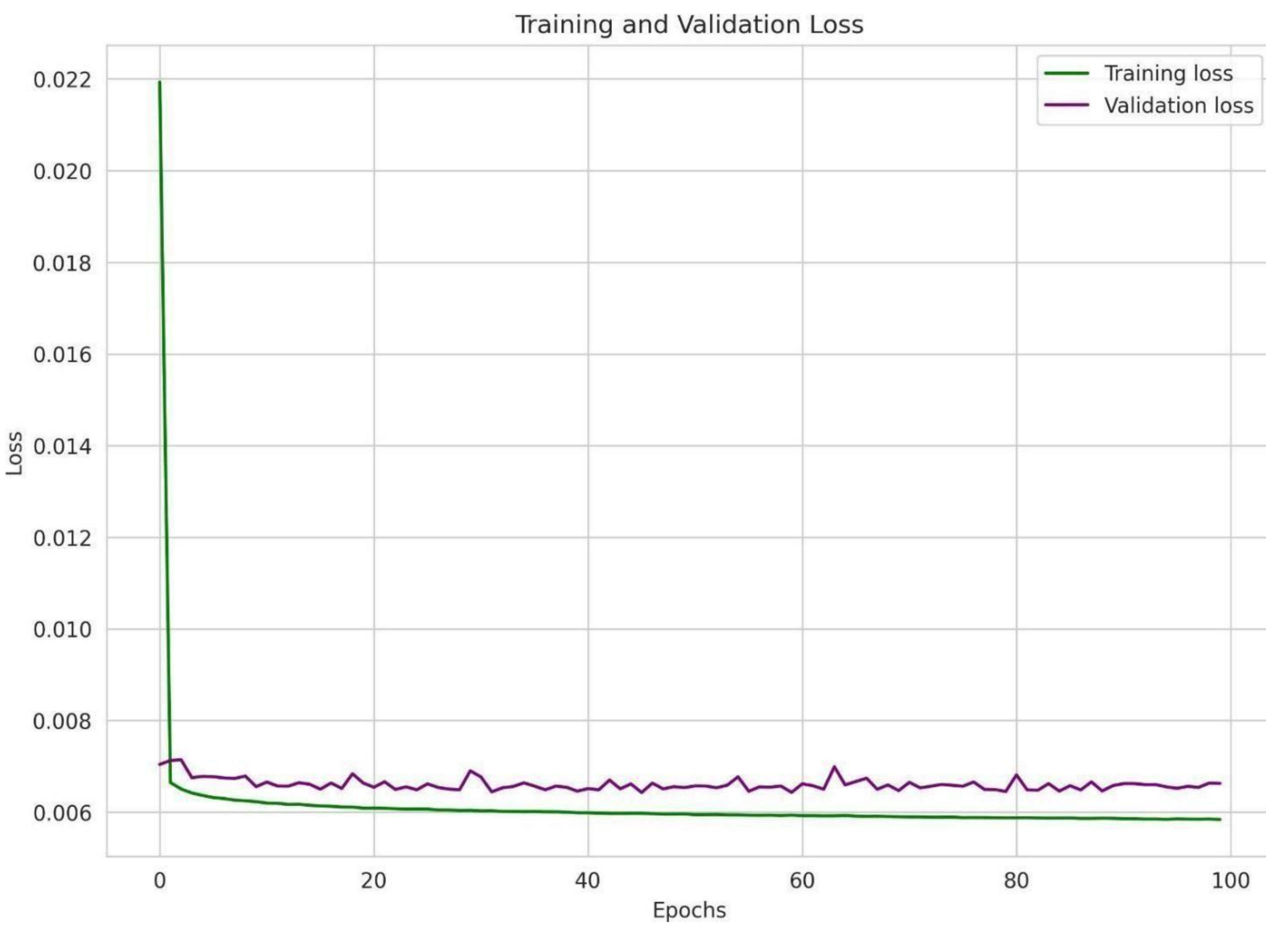

Fig 5.1 Validation and Training loss for RNN

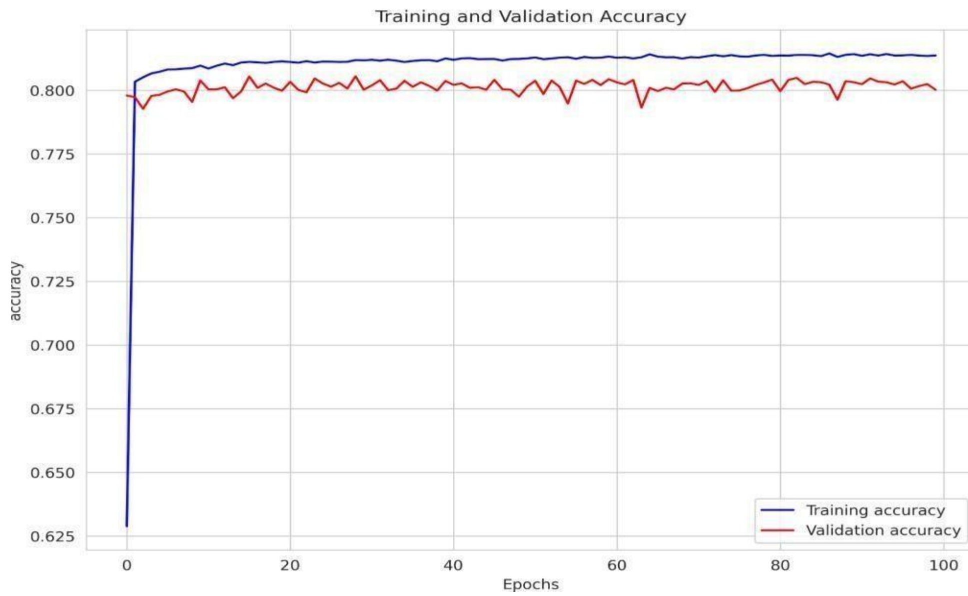

Fig 5.2 Validation and Training Accuracy for RNN

GRU Model:

```
inputs = Input(shape=(timesteps, input_dim))
gru_out = GRU(128, activation='relu', return_sequences=False)(inputs)


dense = Dense(units=128, activation='relu')(gru_out)

output = Dense(units=output_dim, activation='relu')(dense)

gru = Model(inputs=inputs,  outputs=output)


gru.compile(loss='mse', optimizer=keras.optimizers.Adam(
    learning_rate=learning_rate), metrics=['accuracy'])

# plot_model(model, to_file="Bi-LSTM-MultiheadAtt.png",
#           dpi=300, show_shapes=True)

path_checkpoint = "model_checkpoint_MultiAttLSTM_aqiPre.h5"
es_callback = keras.callbacks.EarlyStopping(
    monitor="val_loss", min_delta=0,
    patience=10
)

modelckpt_callback = keras.callbacks.ModelCheckpoint(
    monitor="val_loss",
```

```
    filepath=path_checkpoint,
    verbose=0,
    save_weights_only=True,
    save_best_only=True,
)

# fit network
history = gru.fit(train_X, train_y, epochs=epochs, batch_size=batch_size, validation_data=(valid_X, valid_y),
# callbacks=[es_callback, modelckpt_callback]
                    )
visualize_loss(history, "Training and Validation Loss")
visualize_acc(history, "Training and Validation Accuracy")

# evaluate(model)
```

In Fig 5.3 Training and validation loss graph for GRU model is plotted. These graphs help in the model improvement by showing convergence and detecting overfitting, ensuring optimal performance and generalization. Fig 5.4 the graph for validation and accuracy training for GRU.

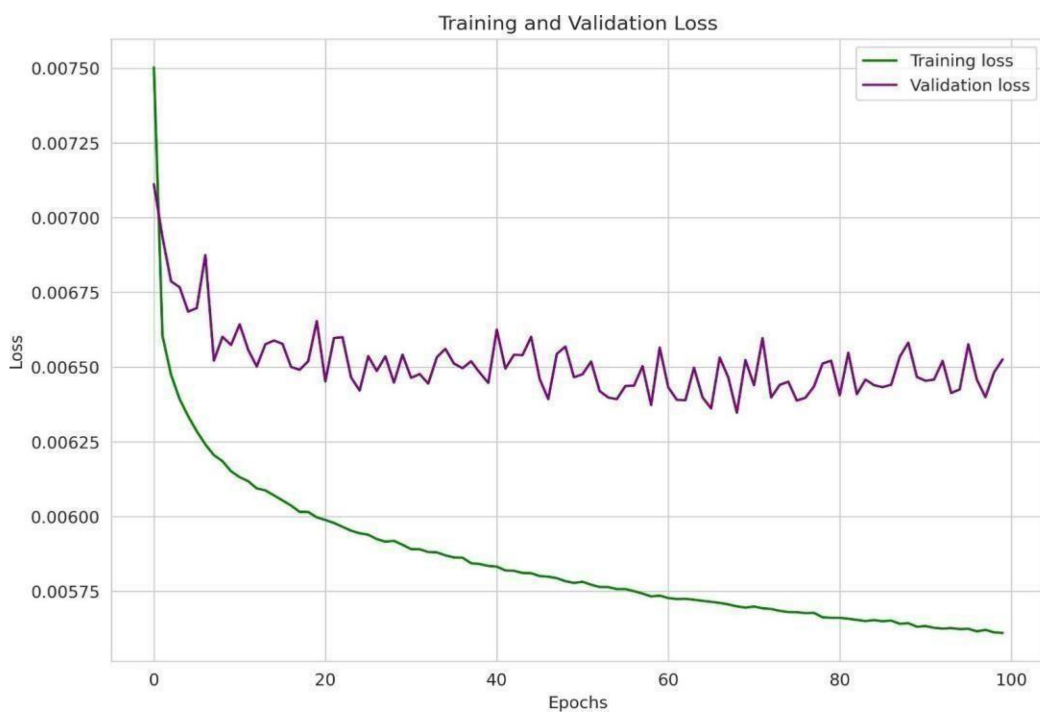

Fig 5.3 Validation and Training loss for GRU

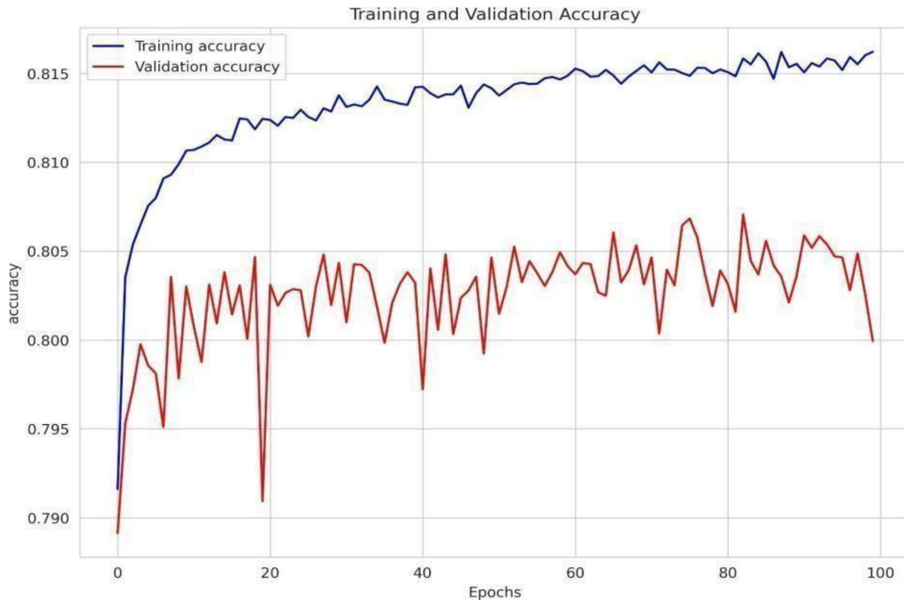

Fig 5.4 Validation and Training accuracy for GRU

LSTM:

```
inputs = Input(shape=(timesteps, input_dim))

lstm_out = LSTM(128, activation='relu', return_sequences=False)(inputs)
dense = Dense(units=128, activation='relu')(lstm_out)

output = Dense(units=output_dim, activation='relu')(dense)

lstm = Model(inputs=inputs,  outputs=output)


lstm.compile(loss='mse', optimizer=keras.optimizers.Adam(
    learning_rate=learning_rate), metrics=['accuracy'])

# plot_model(model, to_file="Bi-LSTM-MultiheadAtt.png",
#           dpi=300, show_shapes=True)

path_checkpoint = "model_checkpoint_MultiAttLSTM_aqiPre.h5"
es_callback = keras.callbacks.EarlyStopping(
    monitor="val_loss", min_delta=0,
    patience=10
)

modelckpt_callback = keras.callbacks.ModelCheckpoint(
```

```
    monitor="val_loss",
    filepath=path_checkpoint,
    verbose=0,
    save_weights_only=True,
    save_best_only=True,
)

# fit network
history = lstm.fit(train_X, train_y, epochs=epochs, batch_size=batch_size, validation_data=(valid_X, valid_y),
#                   callbacks=[es_callback, modelckpt_callback]
                   )
visualize_loss(history, "Training and Validation Loss")
visualize_acc(history, "Training and Validation Accuracy")

# evaluate(model)
```

In Fig 5.5 Training and validation loss graph for LSTM model is plotted. These graphs help in the model improvement by showing convergence and detecting overfitting, ensuring optimal performance and generalization. Fig 5.6 the graph for validation and accuracy training for LSTM.

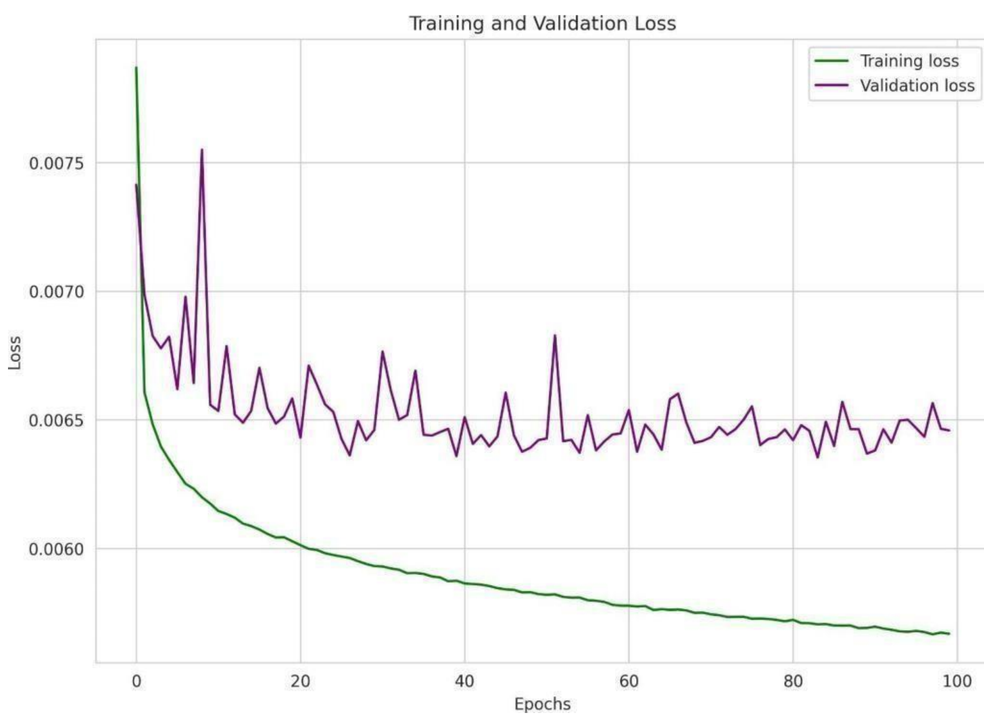

Fig 5.5 Validation and Training loss for LSTM

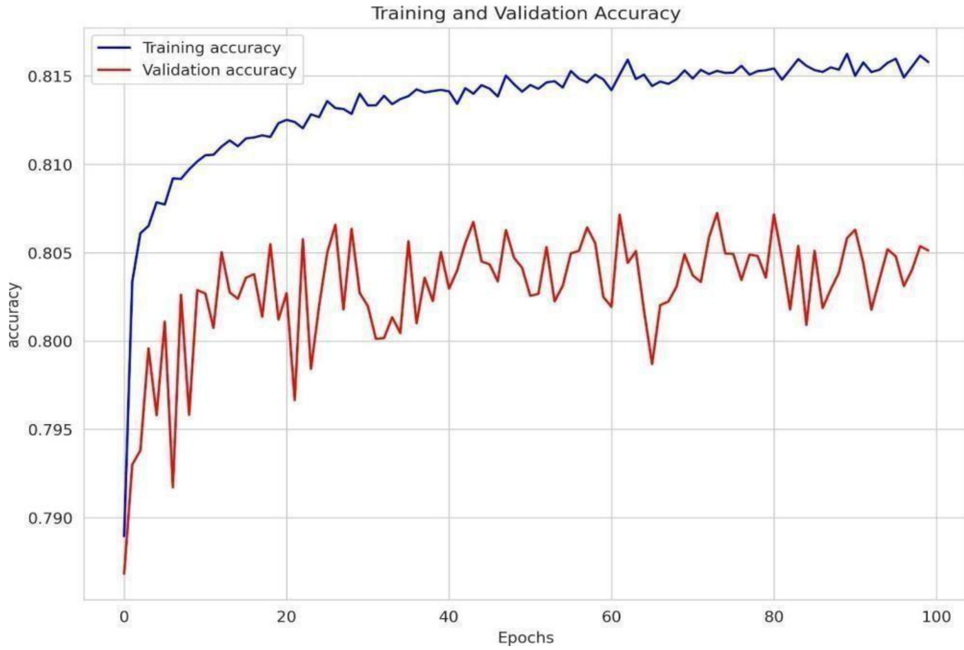

Fig 5.6 Validation and Training Accuracy for LSTM

BiLSTM:

```
inputs = Input(shape=(timesteps, input_dim))

lstm_out = Bidirectional(
    LSTM(units=128, activation='relu', return_sequences=True))(inputs)
multihead_attn = MultiHeadAttention(
    num_heads=4, key_dim=32)(lstm_out, lstm_out)

flatten = Flatten()(multihead_attn)
dense = Dense(units=128, activation='relu')(flatten)

output = Dense(units=output_dim, activation='relu')(dense)

BiLSTMultiheadAtt = Model(inputs=inputs,  outputs=output)


BiLSTMultiheadAtt.compile(loss='mse', optimizer=keras.optimizers.Adam(
    learning_rate=learning_rate), metrics=['accuracy'])

# plot_model(model, to_file="Bi-LSTM-MultiheadAtt.png",
#           dpi=300, show_shapes=True)

path_checkpoint = "model_checkpoint_MultiAttLSTM_aqiPre.h5"
es_callback = keras.callbacks.EarlyStopping(
```

```
    monitor="val_loss", min_delta=0,
    patience=10
)
modelckpt_callback = keras.callbacks.ModelCheckpoint(
    monitor="val_loss",
    filepath=path_checkpoint,
    verbose=0,
    save_weights_only=True,
    save_best_only=True,
)
# fit network
history = BiLSTMMultiheadAtt.fit(train_X, train_y, epochs=epochs, batch_size=batch_size, validation_data=(valid_X,
valid_y),
#                    callbacks=[es_callback, modelckpt_callback]
                   )
visualize_loss(history, "Training and Validation Loss")
visualize_acc(history, "Training and Validation Accuracy")

# evaluate(model)
```

In Fig 5.7 Training and validation loss graph for BiLSTM model is plotted. These graphs help the model improvement by showing convergence and detecting overfitting, ensuring optimal performance and generalization. Fig 5.8 the graph for validation and accuracy training for BiLSTM.

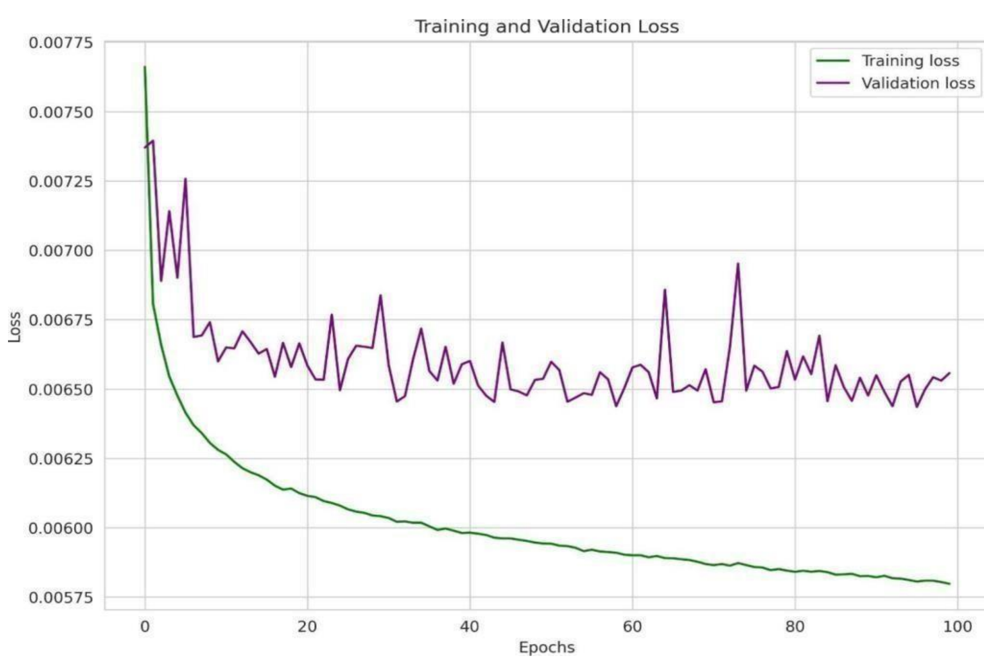

Fig 5.7 Validation and Training loss for BiLSTM

Fig 5.8 Validation and Training Accuracy for BiLSTM

CHAPTER 6

RESULTS AND DISCUSSION

This project involves taking of various deep learning models and using them for the prediction of air quality by considering some significant pollutants that are causing various health diseases and impacting the environment. We used ability of the model to predict accurately, interpret, and maintain a focus on air quality situations. Deep learning's primary benefit is its capacity to identify geographical and temporal relationships in air quality data, which offers a more thorough and precise understanding of air quality dynamics. We overcame the drawbacks of traditional AQI prediction methodologies by utilizing deep learning techniques. This could lead to the development of more potent early warning systems, decision support tools, and enhanced air quality management.

Predicting air quality is a crucial task with several obstacles. The intricate interaction of several variables that affect air quality, such as weather, emissions from different sources, and chemical reactions in the atmosphere, is one of the main challenges. Precise forecasting is a difficult endeavour since these interactions are dynamic and complex. Furthermore, since real-time and spatially distributed data are crucial for predictive model performance, obtaining and maintaining high-quality data can be quite difficult. The complexity is further increased by the introduction of new pollutants, such as volatile organic compounds and fine particulate matter, and the requirement for fine-grained spatial projections. Forecasting air quality is complicated and faced with a number of issues due to geographic factors. Mountains, valleys, and bodies of water are examples of different topography that can have a big impact on local wind patterns and air pollution dispersal. It can be challenging to generalize predictions across wider regions due to the possibility of microclimates with unique air quality characteristics resulting from these topography variances. Forecasting can also be made more difficult by the proximity of emission sources, such as industrial sites or urban areas, to topographical features, which can result in the creation of localized pollution hotspots. Topography can have an impact on meteorological conditions, which can change air stability and mixing and affect how pollutants are transported and dispersed.

6.1 Time Series Prediction

Time series data is essential for air quality prediction because it provides insightful information and useful applications for managing air quality. Time series data have several uses in the forecast of air quality. In order to protect the environment and public health, it is crucial for comprehending the temporal dynamics of air quality, locating pollution sources, and putting into practice efficient remedies. In addition, early warning systems, regulatory compliance, and well-informed policy decisions are all supported by time series analysis and are essential to the continuous fight against air pollution and the negative effects it has on environment and humans.

In this study, the time-series prediction of pollution prediction by the four models RNN,GRU,LSTM and BiLSTM is shown below:

1.) Recurrent Neural Network (RNN)

In (Fig 6.1.1), a time series prediction is done by the RNN model. The graph depicts the actual concentration of the pollutants vs the predicted concentration of pollutants.

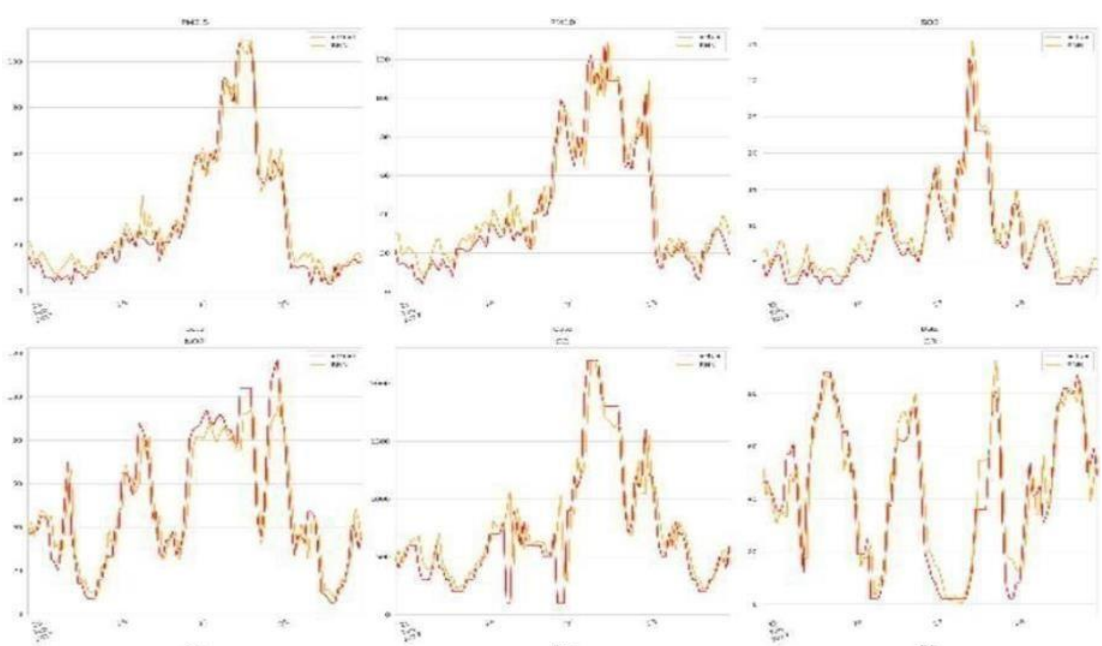

Fig 6.1.1 Time series prediction by RNN

2.) Long Short-Term Memory (LSTM)

In (Fig 6.1.2), a time series prediction is done by the LSTM model. The graph depicts the actual concentration of the pollutants vs the predicted concentration of pollutants.

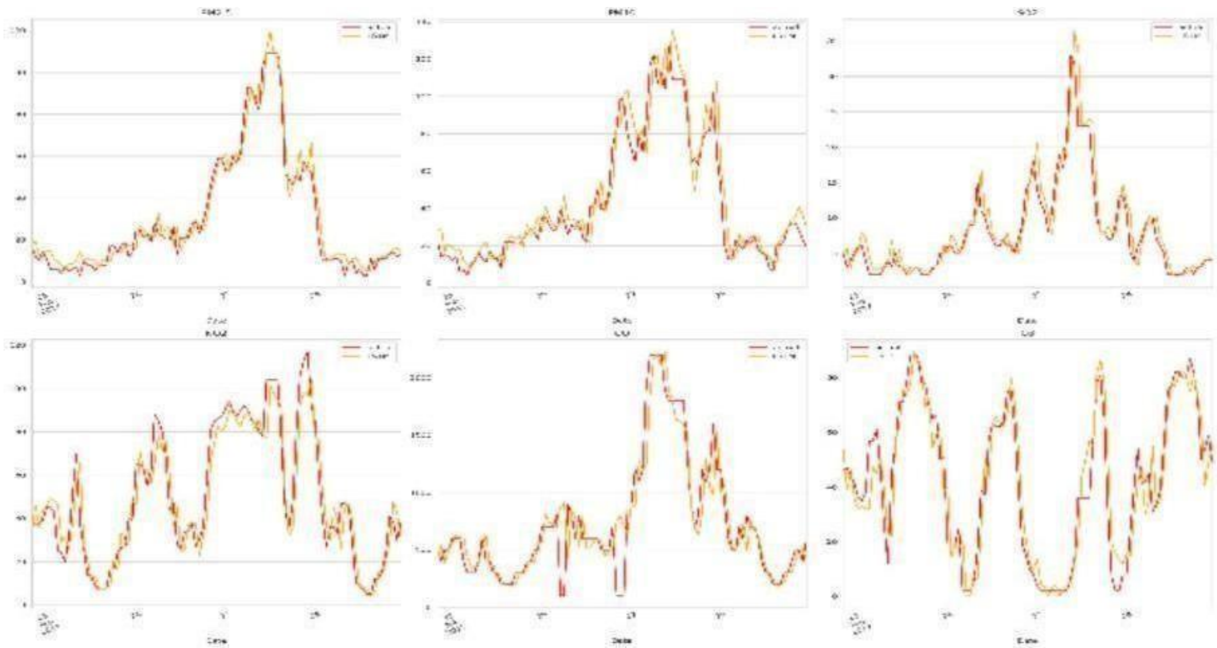

Fig 6.1.2 Time series prediction by LSTM

‘

3.) Gated Recurrent Network (GRU)

In (Fig 6.1.3), a time series prediction is done by the GRU model. The graph depicts the actual concentration of the pollutants vs the predicted concentration of pollutants.

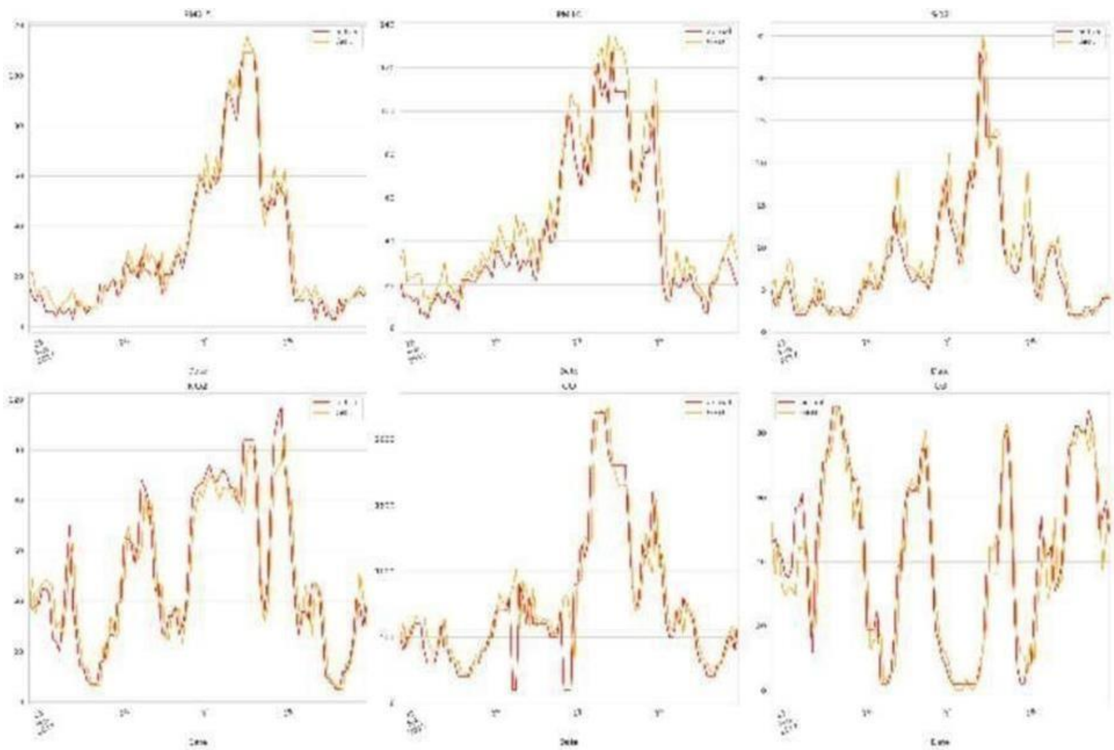

Fig 6.1.2 Time series prediction by GRU

4)Bi directional LSTM

In (Fig 6.1.4), a time series prediction is done by the BiLSTM model. The graph depicts the actual concentration of the pollutants vs the predicted concentration of pollutants.

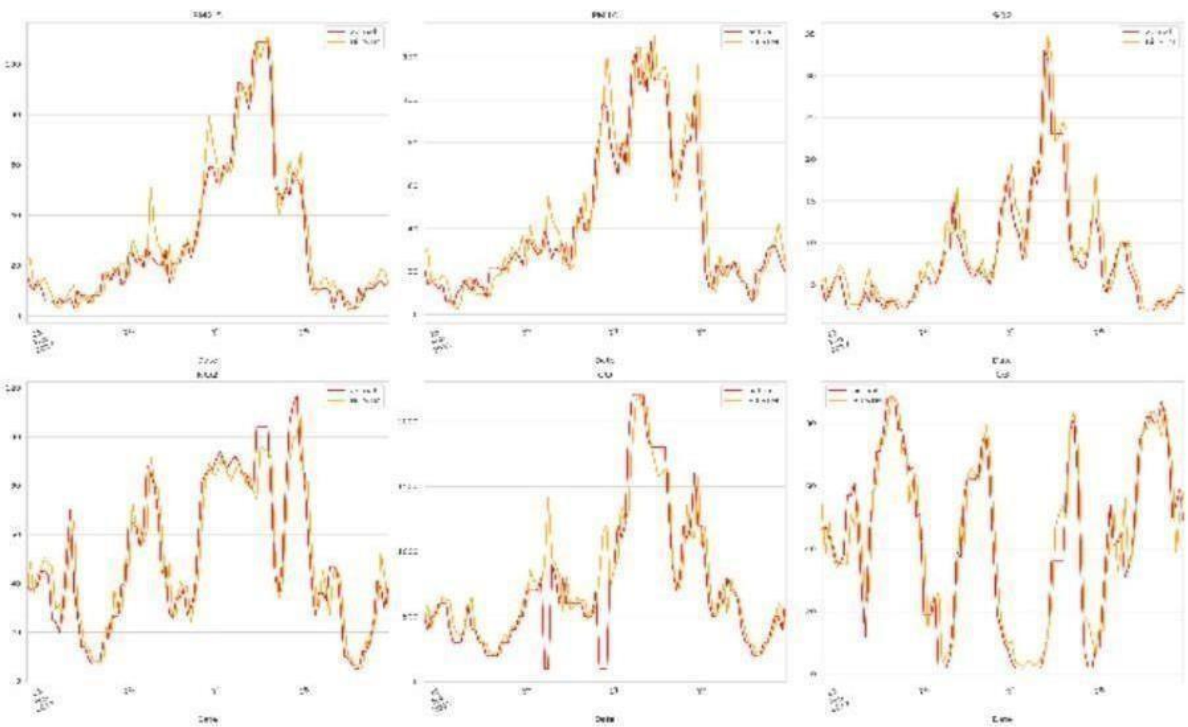

Fig 6.1.4 Time series prediction by Bi LSTM

6.2 Seasonality Trends-

Seasonal patterns in air quality refer to regular fluctuations in the concentrations of pollutants that occur according to a specific timetable all year long. These patterns have important ramifications for the forecast and management of air quality and are impacted by a variety of factors, such as weather, human activity, and natural phenomena. Seasonal trends are calculated for six pollutants namely O3(Fig 6.2.1), NO2(Fig 6.2.2), CO(Fig 6.2.3), SO2(Fig 6.2.4), PM2.5 (Fig 6.2.5) and PM10(Fig 6.2.6).Graphs are plotted for each of the pollutant taking all months in several years period in the X-axis and the concentration of the pollutant in the Y-axis.

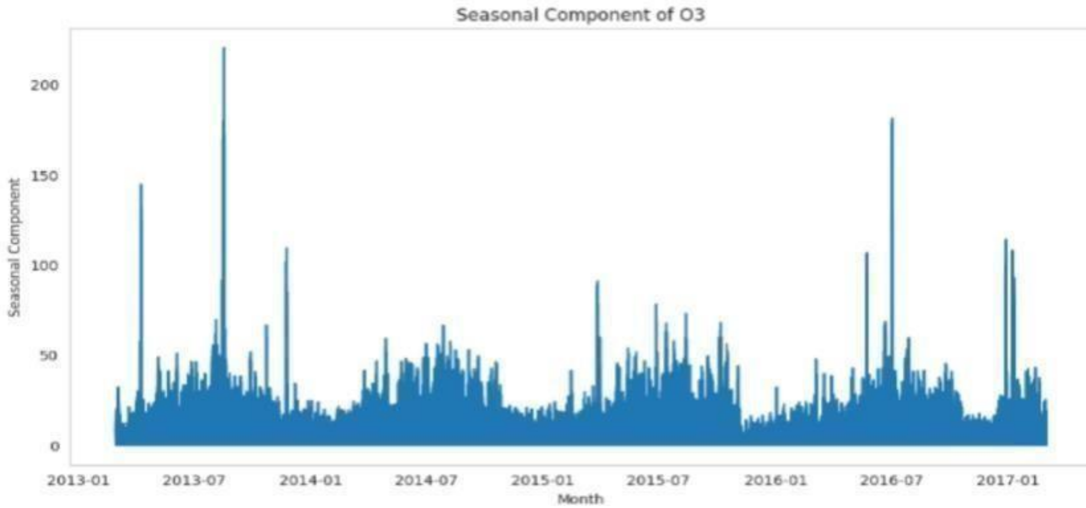

Fig 6.2.1 Seasonal Component of Ozone (O3)

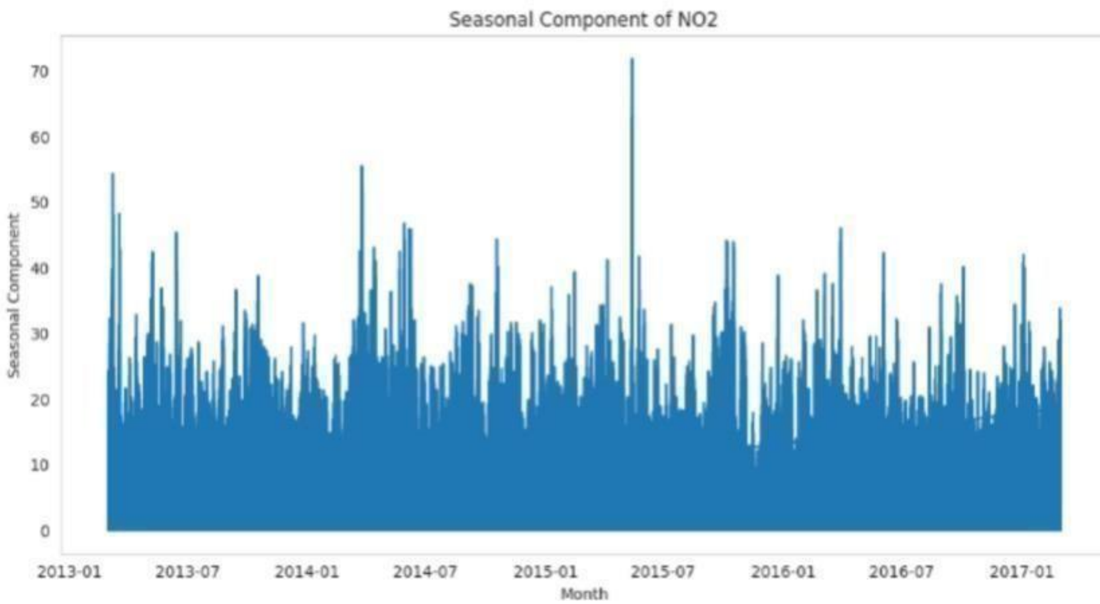

Fig 6.2.2 Seasonal Component of Nitrogen Dioxide (NO2)

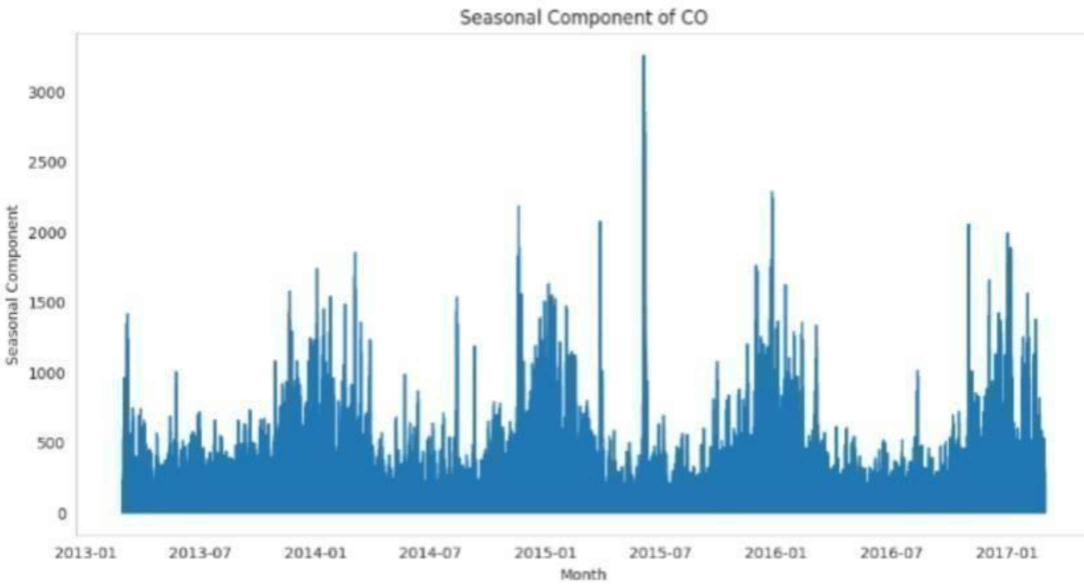

Fig 6.2.3 Seasonal Component of Carbon Mono Oxide (CO)

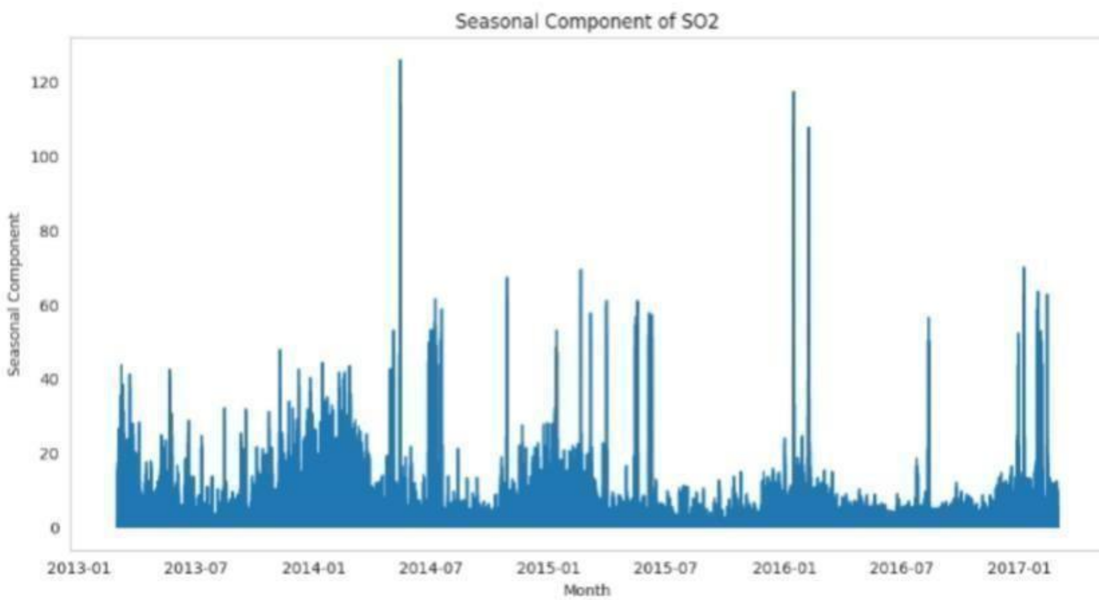

Fig 6.2.4 Seasonal Component of Sulphur Dioxide (SO2)

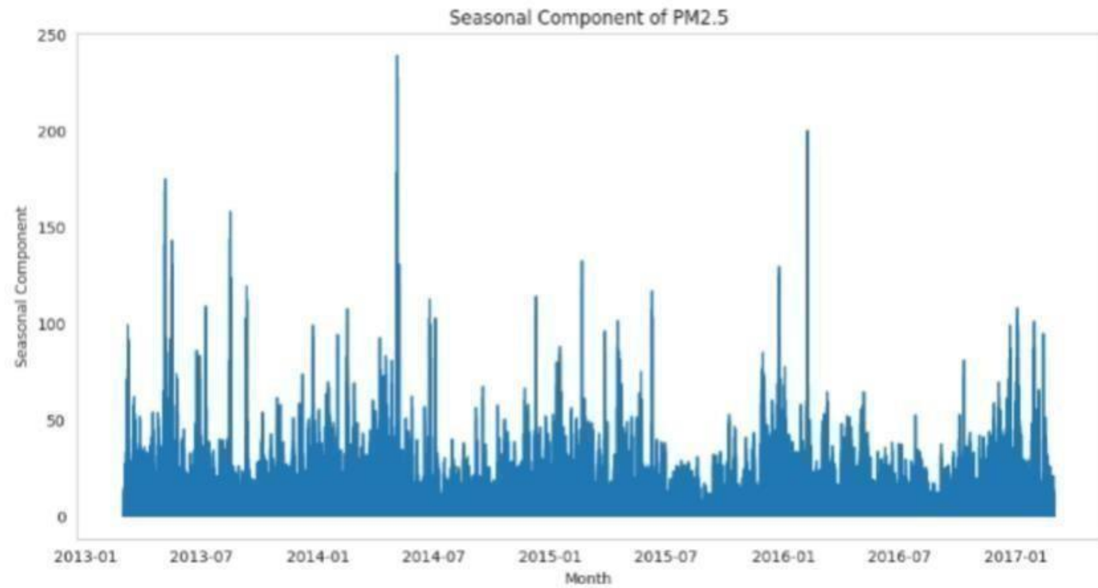

Fig 6.2.5 Seasonal Component of PM2.5

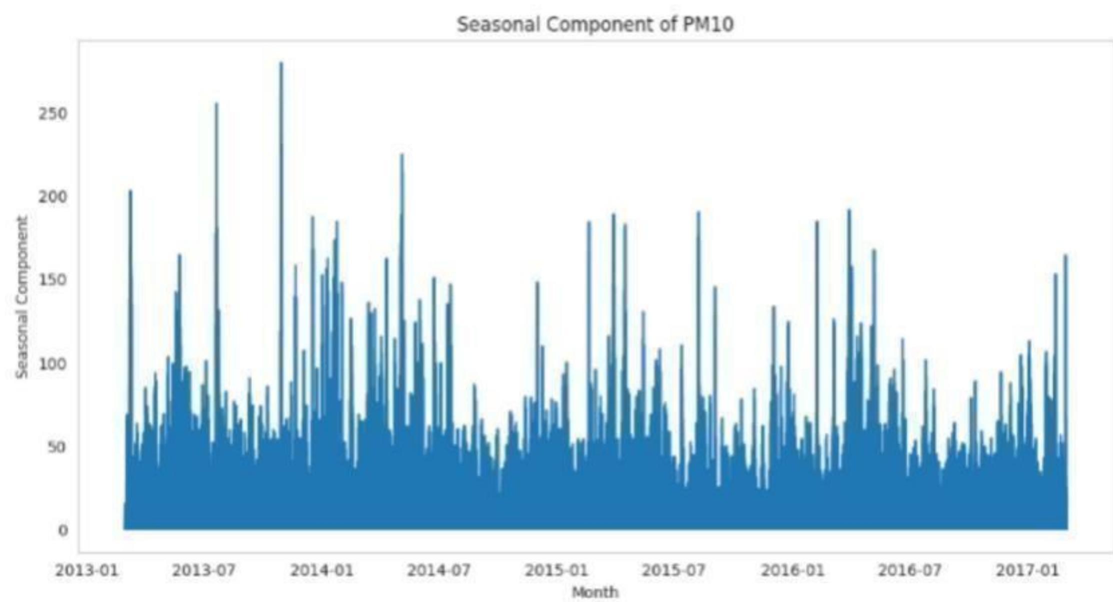

Fig 6.2.6 Seasonal Component of PM10

6.3 Evaluation Matics

Root Mean Square Error

- GRU has the lowest RMSE (101.393), indicating better overall prediction accuracy as shown in Table 6.3 than RNN (102.433) and LSTM (102.745) as indicated in Table 6.3.
- Bi-LSTM-MultiheadAtt also performs well with an RMSE of 101.788, close to GRU.

Mean Absolute Error

- Similar to RMSE, GRU has the lowest MAE (31.690), followed by Bi-LSTM-MultiheadAtt (31.605), LSTM (31.978), and RNN (32.139). Lower MAE values suggest better model performance.

Coefficient of Determination

- R2 measures the proportion of the variance in the dependent variable that is predictable from the independent variables.
- LSTM has the highest R^2 (0.859), followed closely by GRU (0.858) and Bi-LSTM-MultiheadAtt (0.858).
- RNN has a slightly lower R^2 value of 0.856.

Metrics / Models	**RMSE**	**MAE**	**R^2**
RNN	102.4	32.139	0.856
GRU	101.393	31.690	0.858
LSTM	102.745	31.978	0.859
Bi LSTM	101.1	31.605	0.858

CHAPTER 7

CONCLUSION

This time series prediction is a crucial and multidimensional endeavor with significant implications for public health, environmental conservation, and policymaking. This complex task requires integration of deep learning techniques and real-time monitoring. In our proposed approach the prediction of the pollution of the major pollutants is taken into consideration where the model entertains to be trained and tested on today's real-time data. In the future, the model can developed for Health Impact Assessment where the model can assess the potential treat and alert the population regarding the treat in the environment. In order to give the public and local authorities accurate and timely air quality information, we developed an air quality prediction system in this study using deep learning models.

The following are the main findings and observations that we have reached after analyzing both historical and current data:

- Model Performance
- Real Time Prediction
- Spatial Interpolation
- Data Fusion
- Temporal Trends

In conclusion, applying deep learning models to forecast air quality has shown a great deal of promise and potential in addressing the pressing issue of air pollution. Compared to more conventional statistical techniques, these models have shown to be successful in identifying intricate patterns and correlations within air quality data, enabling more precise forecasts. It's important to recognize that there are still difficulties with using deep learning models to predict air quality. Among the main obstacles that must be overcome are those related to data quality, model interpretability, and computational resources. Deep learning models are becoming a more valuable weapon in the fight against air pollution and the negative effects it has on human health and the environment, however, as a result of continued research and development in this area.

CHAPTER 8
FUTURE ENHANCEMENT

While this AQI prediction system has achieved notable success, there are several areas where further developments and enhancements can be explored:

- Improved Data Sources: Understanding the socio-environmental dynamics influencing air quality can also be aided by the incorporation of socio-economic and land-use data. Improved data sources will allow models to capture a more complete picture of the variables affecting air quality, which will result in more precise forecasts.
- Model Optimization: Deep learning models require constant tuning and improvement. This entails enhancing model architectures, adjusting hyperparameters, and investigating cutting-edge methods such as self-attention mechanisms and transfer learning. Prioritizing model efficiency and interpretability can enhance the models' practicality and accessibility for real-world applications.
- Multi Model Data: Predicting air quality more reliably and robustly can be achieved by combining the outputs of several models, such as neural networks, statistical models, and physical models. Better overall performance can be achieved using ensemble techniques and hybrid models that combine the best features of many models, especially in areas with complicated atmospheric circumstances.
- Health Impact Assessment: It is essential to include health effect assessments in the scope of air quality forecast. This entails calculating the health risks connected to particular air contaminants in order to facilitate more focused treatments and public health initiatives. Policymakers can make more informed decisions by gaining insights into the short- and long-term impacts of air pollution on health through the integration of epidemiological and air quality data.
- Public Engagement: It is essential to involve the public in the prediction and control of air quality in order to raise awareness and promote behavioral changes. Creating

easily navigable, real-time applications for air quality prediction and educational campaigns can enable people to take preventative action to save their health.

In terms of providing precise and understandable air quality information, this air quality prediction system has evolved based on the requirements and necessities. Its potential and influence on environmental awareness, public health, and policy decision-making can be further enhanced by tackling the areas that have been identified for future development. This system is going to be crucial in the ongoing effort to create a healthier and cleaner environment and cleaner air.

CHAPTER 9

REFERENCES

[1] Y. Jiao, Z. Wang and Y. Zhang, "Prediction of Air Quality Index Based on LSTM," 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 2019, pp. 17-20, doi: 10.1109/ITAIC.2019.8785602.

[2] C. Zhan, S. Li, J. Li, Y. Guo, Q. Wen and W. Wen, "Prediction of Air Quality in Major Cities of China by Deep Learning," 2020 16th International Conference on Computational Intelligence and Security(CIS), Guangxi, China, 2020, pp. 68-72, doi: 10.1109/CIS52066.2020.00023.

[3] R. Bao, Y. Zhou and W. Jiang, "FL-CNN-LSTM: Indoor Air Quality Prediction Using Fuzzy Logic andCNN-LSTM Model," 2022 2nd International Conference on Electrical Engineering and Control Science (IC2ECS), Nanjing, China, 2022, pp. 986-989, doi: 10.1109/IC2ECS57645.2022.10088050.

[4] K. Manikandan, R. Rajkumar, P. Jayanthi, S. Jothimani, R. Venkatesh and S. Govinda Rao, "An Intelligent Bayesian Optimization with Stacked BiLSTM Model for Air Quality Index Prediction," 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2023, pp. 1699-1704, doi: 10.1109/ICIRCA57980.2023.10220650.

[5] D. Ao, Z. Cui and D. Gu, "Hybrid model of Air Quality Prediction Using K-Means Clustering and Deep Neural Network," 2019 Chinese Control Conference (CCC), Guangzhou, China, 2019, pp. 8416-8421, doi: 10.23919/ChiCC.2019.8865861.

[6] K. M. O. V. K. Kekulanadara, B. T. G. S. Kumara and B. Kuhaneswaran, "Comparative Analysis of Machine Learning Algorithms for Predicting Air Quality Index," 2021 From Innovation To Impact (FITI), Colombo, Sri Lanka, 2021, pp. 1-5, doi: 10.1109/FITI54902.2021.9833033.

[7] J. -C. Cheng and H. -C. Peng, "Air Quality Forecast and Evaluation Based on Long Short-Term Memory Network and Fuzzy Algorithm," 2021 IEEE 4th International Conference on Knowledge Innovation and Invention (ICKII), Taichung, Taiwan, 2021, pp. 88-91, doi: 10.1109/ICKII51822.2021.9574707.

[8] V. Gupta, S. Kapadia and C. Bhadane, "Time Series Analysis and Forecasting of Air Quality in India," 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT), Erode, India, 2023, pp. 1-5, doi: 10.1109/ICECCT56650.2023.10179673.

[9] W. Wang and S. Yang, "Research on Air Quality Forecasting Based on Big Data and Neural

a. Network," 2020 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 2020, pp. 180-184, doi: 10.1109/ICCNEA50255.2020.00045.

[10] C. R. K, N. R. K, B. P. K and P. S. Rajendran, "The Prediction of Quality of the Air Using Supervised Learning," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 2021, pp. 1-5, doi: 10.1109/ICCES51350.2021.9488983.

[11] A. Akanksha, N. Maurya, M. Jain and S. Arya, "Prediction And Analysis of Air Pollution Using Machine Learning Algorithms," 2023 3rd International Conference on Intelligent Technologies (CONIT), Hubli, India, 2023, pp. 1-6, doi: 10.1109/CONIT59222.2023.10205615.

[12] A. Catovic, E. Kadusic, C. Ruland, N. Zivic and N. Hadzajlic, "Air pollution prediction and warning system using IoT and machine learning," 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Maldives, Maldives, 2022, pp. 1-4, doi: 10.1109/ICECCME55909.2022.9987957.

[13] S. Samadi and A. K. Kumawat, "IoT Enabled Low-Cost Real-Time Remote Air Quality Monitoring and Forecasting System," 2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT), Dehradun, India, 2023, pp. 253-258, doi: 10.1109/DICCT56244.2023.10110229.

[14] Y. Song, H. Mao and H. Li, "Spatio-Temporal Modeling For Air Quality Prediction Based On Spectral Graph Convolutional Network And Attention Mechanism," 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 2022, pp. 1-9, doi: 10.1109/IJCNN55064.2022.9891881.

[15] A. Bekkar, B. Hssina, S. Douzi and K. Douzi, "Air Quality Forecasting using decision trees algorithms," 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, 2022, pp. 1-4, doi: 10.1109/IRASET52964.2022.9737814.

[16] Y. Zhongjie, W. Shengwei and W. Ze, "Air quality prediction method based on the CS-LSTM," 2022 5th International Conference on Data Science and Information Technology (DSIT), Shanghai, China, 2022, pp. 1-5, doi: 10.1109/DSIT55514.2022.9943922.

[17] W. Shi and A. Song, "Air Quality Prediction Model Based on Spatio-temporal Graph Convolution Neural Networks," 2022 4th International Academic Exchange Conference on Science and Technology Innovation (IAECST), Guangzhou, China, 2022, pp. 1015-1018, doi:

a. 10.1109/IAECST57965.2022.10062144.

[18] Yue-Shan Chang, Hsin-Ta Chiao, Satheesh Abimannan, Yo-Ping Huang, Yi-Ting Tsai, Kuan-Ming Lin,An LSTM-based aggregated model for air pollution forecasting, AtmosphericPollution Research, Volume 11, Issue 8,2020,Pages 1451-1463,ISSN 1309-

a. 1042,https://doi.org/10.1016/j.apr.2020.05.015.

[19] Luo Zhang, Peng Liu, Lei Zhao, Guizhou Wang, Wangfeng Zhang, Jianbo Liu,Air quality predictions with a semi-supervised bidirectional LSTM neural network,Atmospheric Pollution Research,Volume 12, Issue 1,2021,Pages 328-339,ISSN 1309-1042, https://doi.org/10.1016/j.apr.2020.09.003.

[20] M. Ramezani, H. Habibi, J. L. Sanchez-Lopez and H. Voos, "UAV Path Planning Employing MPC-Reinforcement Learning Method Considering Collision Avoidance," 2023 International Conference on Unmanned Aircraft Systems (ICUAS), Warsaw, Poland, 2023, pp. 507-514, doi: 10.1109/ICUAS57906.2023.10156232.

[21] Athira V, Geetha P, Vinayakumar R, Soman K P,DeepAirNet: Applying Recurrent Networks for Air Quality Prediction,Procedia Computer Science,Volume 132,2018,Pages 1394-1403,ISSN 1877-0509,https://doi.org/10.1016/j.procs.2018.05.068.

[22] Dey P, Dev S, Schoen-Phelan B. NeSDeepNet: A Fusion Framework for Multi-step Forecasting of Near-surface Air Pollutants. In2023 Photonics & Electromagnetics Research Symposium (PIERS) 2023 Jul 3 (pp. 1826-1834). IEEE.

[23] L.F. Wu, N. Li and Y.J. Yang, "Prediction of air quality indicators for the Beijing-Tianjin-Hebei region", Journal of Cleaner Production, vol. 196, 2018.

[24] C. Deng, L. Liu, C. Wang and Z. Chen, "Air Quality Prediction Based on Graph Attention Network," 2023 4th International Conference on Electronic Communication and Artificial Intelligence (ICECAI), Guangzhou, China, 2023, pp. 364-369, doi: 10.1109/ICECAI58670.2023.10176911.

[25] C. Li, Y. Li and Y. Bao, "Research on Air Quality Prediction Based on Machine Learning," 2021 2nd International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), Shenyang, China, 2021, pp. 77-81, doi: 10.1109/ICHCI54629.2021.00022.

APPENDIX A
CONFERENCE SUBMISSION

Our paper abstract titled" Air Quality Prediction using deep learning models" was submitted for consideration at the SOCSDG2023. Approval for paper submission has been received. On submitting this paper and presenting it, there may be a chance to get it published.

Acceptance - 2nd International Conference on Higher Education Institute's Challenges & Solutions for Sustainable Development Goals 2023 for SDG9 External Inbox ×

CTECH SRM Wed, Oct 18, 3:23 PM

to aa9959, sg0755, vr9815, pb6225, da8500, me, mr8697, kk4937, ar9117, cq5256, sq9322, uu7865, sp6247, as2240, manojk3, ns0956, s

Dear Author(s),

Warm greetings.

We are pleased to inform you that your paper has been accepted for presentation at the "**2nd International Conference on Higher Education Institute's Challenges & Solutions for Sustainable Development Goals 2023**," which will be held at SRM Institute of Science and Technology, Kattankulathur from November 1, 2023, to November 3, 2023. To proceed with your registration, please make the payment as instructed and refer to the attachment for your "PAPER ID," which is required for filling out the registration form. The detailed schedule for the conference will be shared with you soon.

Kindly fill the below forms. It's **mandatory** to fill both the forms.

FigureA.1: SOCSDG2023 Acceptance

APPENDIX B

PLAGIARISM REPORT

Air Quality_15.11.23

ORIGINALITY REPORT

8%	5%	4%	5%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ijircce.com Internet Source	1%
2	Submitted to University of Southampton Student Paper	1%
3	Submitted to Stamford College Student Paper	1%
4	Submitted to Carrington College Student Paper	1%
5	Manjit Kaur, Dilbag Singh, Mohamed Yaseen Jabarulla, Vijay Kumar, Jusung Kang, Heung-No Lee. "Computational deep air quality prediction techniques: a systematic review", Artificial Intelligence Review, 2023 Publication	<1%
6	Submitted to University of Strathclyde Student Paper	<1%
7	Rekha Rajagopal. "Comparative Analysis of COVID-19 X-ray Images Classification Using Convolutional Neural Network, Transfer Learning, and Machine Learning Classifiers	<1%

Using Deep Features", Pattern Recognition and Image Analysis, 2021
Publication

8	jewlscholar.mtsu.edu Internet Source	<1 %
9	www.grin.com Internet Source	<1 %
10	Submitted to King's College Student Paper	<1 %
11	www.frontiersin.org Internet Source	<1 %
12	www.mdpi.com Internet Source	<1 %
13	Isaiah Hull. "Machine Learning for Economics and Finance in TensorFlow 2", Springer Science and Business Media LLC, 2021 Publication	<1 %
14	"Natural Language Processing and Chinese Computing", Springer Science and Business Media LLC, 2018 Publication	<1 %
15	Submitted to National College of Ireland Student Paper	<1 %
16	Submitted to Rochester Institute of Technology Student Paper	<1 %

17	riunet.upv.es Internet Source	<1 %
18	Submitted to Kingston University Student Paper	<1 %
19	Submitted to SRM University Student Paper	<1 %
20	Saeed Mohsen, Anas M. Ali, Ahmed Emam. "Automatic modulation recognition using CNN deep learning models", Multimedia Tools and Applications, 2023 Publication	<1 %
21	medium.datadriveninvestor.com Internet Source	<1 %
22	www.cnblogs.com Internet Source	<1 %
23	grotrian.nsu.ru Internet Source	<1 %
24	mdpi-res.com Internet Source	<1 %
25	ntnuopen.ntnu.no Internet Source	<1 %

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Deemed to be University u/s 3 of UGC Act, 1956)

Office of Controller of Examinations

REPORT FOR PLAGIARISM CHECK ON THE SYNOPSIS/THESIS/DISSERTATION/PROJECT REPORTS

1	Name of the Candidate **(IN BLOCK LETTERS)**	SASHANK DONVALLI HARSHITHA KAMBHAM
2	Address of the Candidate	SRM University Chennai **Mobile Number :** 9347270682,9515960651
3	Registration Number	RA2011026010234 RA2011026010235
4	Date of Birth	21/09/2002 02/06/2002
5	Department	Computational Intelligence(CINTEL)
6	Faculty	Faculty of Engineering and Technology
7	Title of the Synopsis/ Thesis/ Dissertation/Project	Air Quality Prediction using Deep Learning Models
8	Name and address of the Supervisor / Guide	Dr. M. Dhilsath Fathima GUIDE ASSISTANT PROFESSOR Dept of Computational Intelligence **Mail ID : dhilsatm@srmist.edu.in** **Mobile Number :8668122854**
9	Name and address of the Co-Supervisor / Co- Guide (if any)	**Mail ID :** **Mobile Number :**
10	Software Used	Turnitin

Date :

11	Date of Verfication			
12	**Plagiarism Details: (to attach the final report)**			
Chapter	**Title of the Chapter**	**Percentage of similarity index (including self citation)**	**Percentage of similarity index (Excluding self citation)**	**% of plagarism after excluding Quotes, Bibliography, etc.,**
1	Introduction	2	1.5	0
2	Literature Survey	1	0	0
3	System Architecture and Design	1	0	0
4	Methodology	1.5	0	0
5	Coding and Testing	1.5	0.5	0
6	Results and Discussion	0	0	0
7	Conclusion	0	0	0
8	Future Enhancement	0	0	0
9				
10				
Thesis abstract		1	0	0
Appendices		0	0	0

I / We declare that the above information have been verified and found true to the best of my / our knowledge.

Signature of the Candidate	**Signature of the Supervisor / Guide**
Signature of the Co-Supervisor/Co-Guide	**Signature of the HOD**

Date :