Impact of COVID-19 lockdown on air quality analyzed through machine learning techniques

Umer Zukaib [1, 2] , **Mohammed Maray** [3] , **Saad Mustafa** [Corresp., 1] , **Nuhman ul Haq** [1] , **Atta ur Rehman Khan** [4] , **Faisal Rehman** [1]

[1] Computer Science, COMSATS University Islamabad, Abbottabad Campus, Abbottabad, KP, Pakistan

[2] Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, Wuhan, China

[3] College of Computer Science and Information Systems, King Khalid University, Abha, Saudi Arabia

[4] College of Engineering and Information Technology, Ajman University, Ajman, UAE

Corresponding Author: Saad Mustafa
Email address: saadmustafa@cuiatd.edu.pk

After February 2020, the majority of the world's governments decided to implement a lockdown in order to limit the spread of the deadly COVID-19 virus. This restriction improved air quality by reducing emissions of particular atmospheric pollutants from industrial and vehicular traffic. In this study, we look at how the COVID-19 shutdown influenced the air quality in Lahore, Pakistan. HAC Agri Limited, Dawn Food Head Office, Phase 8-DHA, and Zeenat Block in Lahore were chosen to give historical data on the concentrations of many pollutants, including PM2.5, PM10 (particulate matter), NO2 (nitrogen dioxide), and O3 (ozone) (ozone). We use a variety of models, including Decision Tree, SVR, Random Forest, ARIMA, CNN, N-BEATS, and LSTM, to compare and forecast air quality. Using machine learning methods, we looked at how each pollutant's levels changed during the lockdown. It has been shown that LSTM estimates the amounts of each pollutant during the lockout more precisely than other models. The results show that during the lockdown, the concentration of atmospheric pollutants decreased, and the air quality index improved by around 20%. The results also show a 42% drop in PM2.5 concentration, a 72% drop in PM10 concentration, a 29% drop in NO2 concentration, and an increase of 20% in O3 concentration. The machine learning models are assessed using the RMSE, MAE, and R-SQUARE values. The LSTM measures NO2 at 4.35%, O3 at 8.2%, PM2.5 at 4.46%, and PM10 at 8.58% in terms of MAE. It is observed that the LSTM model outperformed with the fewest errors when the projected values are compared with the actual values.

1 # Impact of COVID-19 Lockdown on Air Quality
2 # Analyzed through Machine Learning Techniques

4 Umer Zukaib[1,2], Mohammed Maray[3], Saad Mustafa[1], Nuhman ul Haq[1], Atta ur Rehman Khan[4],
5 Faisal Rehman[1]
6
7 [1] Department of Computer Science, COMSATS University Islamabad, Abbottabad Campus,
8 Pakistan
9 [2] Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of
10 Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, China
11 [3] College of Computer Science and Information Systems, King Khalid University, Abha, Saudi
12 Arabia
13 [4] College of Engineering and Information Technology, Ajman University, Ajman, UAE
14
15 Corresponding Author:
16 Saad Mustafa[1]
17 Email address: saadmustafa@cuiatd.edu.pk

19 ## Abstract

20 After February 2020, the majority of the world's governments decided to implement a lockdown
21 in order to limit the spread of the deadly COVID-19 virus. This restriction improved air quality
22 by reducing emissions of particular atmospheric pollutants from industrial and vehicular traffic.
23 In this study, we look at how the COVID-19 shutdown influenced the air quality in Lahore,
24 Pakistan. HAC Agri Limited, Dawn Food Head Office, Phase 8-DHA, and Zeenat Block in
25 Lahore were chosen to give historical data on the concentrations of many pollutants, including
26 PM2.5, PM10 (particulate matter), NO2 (nitrogen dioxide), and O3 (ozone) (ozone). We use a
27 variety of models, including Decision Tree, SVR, Random Forest, ARIMA, CNN, N-BEATS,
28 and LSTM, to compare and forecast air quality. Using machine learning methods, we looked at
29 how each pollutant's levels changed during the lockdown. It has been shown that LSTM
30 estimates the amounts of each pollutant during the lockout more precisely than other models.
31 The results show that during the lockdown, the concentration of atmospheric pollutants
32 decreased, and the air quality index improved by around 20%. The results also show a 42% drop
33 in PM2.5 concentration, a 72% drop in PM10 concentration, a 29% drop in NO2 concentration,
34 and an increase of 20% in O3 concentration. The machine learning models are assessed using the
35 RMSE, MAE, and R-SQUARE values. The LSTM measures NO2 at 4.35%, O3 at 8.2%, PM2.5
36 at 4.46%, and PM10 at 8.58% in terms of MAE. It is observed that the LSTM model
37 outperformed with the fewest errors when the projected values are compared with the actual
38 values.

40

1 Introduction

41

42 The COVID-19 pandemic has produced major problems in health and socioeconomic activity all
43 throughout the world [1, 2, 38]. Certain cofactors, such as hazardous atmospheric pollutants,
44 should be considered in the context of disease spread and increasing fatality rates, particularly in
45 densely populated regions [3, 4]. According to one study, the infection rates of the fatal virus
46 COVID-19 and atmospheric pollutants including PM10, CO, SO2, and NH3 have a direct
47 association; as pollution levels rise, so do COVID-19 infection rates [5, 6]. Lockdown, on the
48 other hand, dramatically improves air quality [7]. According to a Chinese study [8], the
49 concentration of air pollutants reduced by 5 to 24 percent during the lockdown, with the decline
50 reaching 40 percent in megacities. Another study of megacities found that lockout reduces
51 PM2.5 concentrations by 20–60% [9]. Furthermore, another environmental study discovered that
52 the fall in industrial activity and transportation is to blame for the rapid drop in air pollution
53 levels [10, 11].

54

55 Lockdown has been criticized for various economic crises and is regarded to be a barrier to
56 progress. It decreases the transmission of viruses and reduces pollutants, but it has a long-term
57 negative impact on economic growth [12]. In the majority of countries, local authorities imposed
58 the lockdown. Therefore, the research community will now have a once-in-a-lifetime chance to
59 investigate the impacts of lockdown on air quality [13]. Utilizing trajectory-based models, some
60 research investigated the impact of lockdown on the quality of the air [14, 15]. Researchers have
61 also looked at the independent factors that affect air quality using fixed-effect models [16, 17].
62 The majority of nations have seen an improvement in air quality during these lockdown times.
63 The concentrations of several pollutants, including PM2.5, NO2, SO2, and CO2, are noticeably
64 lowered in Wuhan, China, during the lockdown [18]. Various studies have presented air quality
65 prediction and forecasting models both during and after lockdown. However, bidirectional
66 LSTM and encoder-decoder based LSTM with multivariate data showed better results [19].

67

68 In this study, we estimate air quality using decision tree regression, support vector regression,
69 Random Forest regression, auto-regressive integrated moving average (ARIMA), Convolution
70 neural networks (CNN), N-BEATS, and long short-term memory (LSTM). Machine learning
71 approaches are used to investigate the improvement in air quality during the COVID-19
72 shutdown in Lahore, Pakistan. Although all of the models yielded excellent results, the LSTM
73 model performed the best due to its superior multivariate problem-solving capabilities. We
74 examined historical air quality data from four locations in Lahore, Pakistan, namely "HAC Agri
75 Limited," "Dawn Food Head Office," "Phase 8-DHA," and "Zeenat Block," which included
76 average concentrations of several pollutants: PM2.5, PM10 (particulate matter), NO2 (nitrogen
77 dioxide), and O3 (ozone). For the LSTM model to forecast these contaminants' anticipated
78 concentrations during the lockdown period, the average concentration of these pollutants is set as
79 a goal variable. The LSTM model may be generalized well. As a result, LSTM is an appropriate

80   deep learning model for analyzing the decrease in pollutant concentration during the COVID-19
81   lockout. The LSTM model is used to analyze historical air quality data obtained from four
82   locations. LSTM used training data from January 2018 through December 2019. A separate
83   dataset, the validation set, is used for validation. The data is covered from January 2020 to May
84   2020, although the actual lockdown time is from March 2020 to May 2020. The model also
85   projected the four-day moving average for many air pollutants. The results reveal that the LSTM
86   model provides more accurate predictions with fewer errors.
87
88   Major contributions of this article are:
89   • To determine the main traits, we used Principal Component Analysis (PCA) on the AQI
90   dataset.
91   • We used scatter plots and the Mahalanobis Distance (MD) technique to find outliers and
92   impute their values using the column mean.
93   • Several machine learning algorithms were used to forecast the concentration of various
94   pollutants.
95   The rest of the article is organized as follows. Section 2 presents the related work. Section 3
96   shows the material and methods. Section 4 presents the experimentation and results. Section 5
97   presents conclusions and future work.
98

2 Related Work

100   Atmospheric pollutants such carbon monoxide, nitrogen oxide, particulate matter, and others are
101   the primary source of pollution in metropolitan areas. These pollutants have a significant
102   negative effect on human health and can lead to a number of conditions, including respiratory
103   difficulties, immune system problems, and cardiovascular abnormalities [20]. Traffic and
104   industrial activity were reduced during the pandemic COVID-19 shutdown, which also resulted
105   in a drop in the concentration of air pollutants. Pollutant concentrations including NO2, and
106   particulate matter have decreased, according to satellite and ground assessments (PM). The
107   cessation of industrial and vehicular activity was the cause of the decrease in pollutant
108   concentration [21]. The atmospheric composition and air pollution are strongly correlated with
109   climatic variables such as humidity, temperature, and the emission of certain industrial chemicals
110   [22]. A program known as the "sulfur oxide emission control program" reportedly began in Los
111   Angeles in 1978. The ultimate goal was to reduce air pollution, particularly sulfur oxide
112   emissions. There was a considerable reduction in the emission of several pollutants, including
113   sulfur oxides, black carbon, and PM2.5, in the very following year [23].
114
115   When we compare toxic pollutants to the rise in COVID-19 infection rates, we discover that both
116   are exactly related to one another: the greater the pollution concentration, the higher the infection
117   rate [5]. Toxic pollutants in the atmosphere can induce lung disorders. A study found that during
118   the COVID-19 shutdown, the air quality of 44 Chinese cities improved from 5 to 24 percent,
119   with megacities like Sao Paulo witnessing a 70 percent decrease in pollution [24]. According to

120 the report, industrial and traffic-related activities are the biggest contributors to air pollution.
121 Road traffic and industrial activity were greatly diminished as a result of the shutdown. The air
122 quality is improved as a result. Researchers have provided a case study on how the COVID-19
123 partial shutdown has reduced air pollution and enhanced air quality in Brazil in [25].
124 Additionally, they have looked at whether the levels of CO2, PM-10, and nitrogen oxide
125 decreased gradually. However, the partial lockdown has not produced sufficient results in terms
126 of the improvement of the air quality index, thus credit should once again be given to the limited
127 activities associated with traffic and industry.
128

129 The report on the air-quality index during the COVID-19 partial shutdown for the "Yangtze
130 River Delta" region in northern China was provided by the researchers in [26]. The author claims
131 that there hasn't been any improvement in the ozone layer documented, however there was
132 evidence that indicated pollution levels were still high during lockdowns and that the causes
133 were primarily chemical industry. Due to some factories in the northern portion of China still
134 operating, a partial lockdown was enforced on parts of the regions, which is why the air quality
135 index was unsatisfactory in some particular areas. The researcher did a study in China [27] while
136 the country was under lockdown. Drop in transportation activity did not result in a decrease in air
137 pollution; instead, climatic phenomena like temperature changes and moisture patterns had a
138 major role. In addition, the author asserted that the decrease in heat waves, rain, thunderstorms,
139 wind, lack of dust storms, tropical storms, derecho (straight-line wind), and anticyclone were the
140 true causes of the increase in air quality. A link between the COVID-19 infection rate and
141 variations in the weather has been established, according to a study done in Jakarta, Indonesia.
142 Due to various meteorological factors, including temperature, humidity, rainfall, air density, air
143 pressure, and precipitation, the infection rate fluctuates throughout Indonesia.
144

145 In [29, 30], scientists investigated the relationship between the COVID-19 mortality rate and
146 changes in the climatic conditions. In the beginning, they focused on Wuhan, China, and
147 gathered the daily mortality records brought on by COVID-19 infection. The researchers then
148 tracked the fluctuations in air contaminants as well as the change in climatic conditions. Based
149 on statistical study, the correlation was created. The research investigation came to the
150 conclusion that the increased death rate among COVID patients was also caused by the high
151 intensity of humidity and temperature changes. Deep neural networks are regarded as being the
152 most reliable model for the forecasting of air quality and the concentration of certain atmospheric
153 contaminants when compared to the conventional statistical model. While performing
154 forecasting and prediction-related activities, it is crucial to take into account the adaption of
155 suitable models that produce accurate prediction results, even if deep neural networks are
156 expensive in terms of computation, optimal-convergence, inefficient prediction due to noisy data,
157 and issues related to over- and under-fitting [31].
158

159 A pruned-neural-network model that forecasts the concentration of PM10 pollution was put up
160 by researchers in [32]. They conducted future projections based on past predictions using records
161 that were largely from prior years and made up of historical data collected on an hourly basis.
162 The aggregated-LSTM model was employed by researchers in [33] to forecast air quality. The
163 outcomes of LSTM were then contrasted with those of support vector regression and regression-
164 based gradient boosting trees. LSTM outperformed other models in terms of prediction
165 performance. [34] describes how researchers employed several machine-learning models to
166 anticipate air quality by examining PM2.5 concentrations at various sizes. Multiple regression
167 trees, deep neural networks, and hybrid models based on LSTM were the models that were
168 employed. However, LSTM has achieved good outcomes when it comes to forecasting and
169 examining the efficiency results for air quality.

170

171 Various experiments have been conducted to determine how COVID-19 lock-down affects air
172 quality. The study demonstrated how lockdown affected the air quality in six Chinese
173 megacities. The four-month lockout had a tremendously favorable effect on the air quality.
174 Reduced industrial activity and decreased traffic were both factors in the improvement of the air
175 quality, which led to a 37 percent increase in the air quality index as a whole. According to a
176 study that was published in [35], when there is a lockdown, the concentration of Ozone increases
177 noticeably while the concentration of other environmental pollutants decreases. An Indian study
178 revealed that the air quality has improved due to the COVID-19 lock-down. The concentration of
179 PM10 was reduced by 60%, that of PM2.5 by 39%, and the air quality as a whole was improved
180 by 40–50%. According to a study conducted at 126 different metropolitan locations in the UK,
181 the lock-down time saw an improvement in the overall quality of the air, a reduction in NO2
182 concentration of around 48 percent, and an increase in ozone (O3) of about 11 percent.

183

184 ## 3 Materials & Methods

185 The LSTM model is regarded as a robust approach to handle multivariate issues [2]. Using
186 historical data, the LSTM model compares the anticipated true values of air contaminants
187 following lockdowns.

188

189 ### 3.1 Data Description

190 Data on atmospheric pollution for Lahore, Pakistan, was gathered from IQAir.
191 https://www.iqair.com/us/pakistan/punjab/lahore. A firm in Switzerland that specializes in
192 preserving and monitoring airborne contaminants is called IQAir. It also deals with establishing
193 and overseeing air cleansing and quality. IQAir has received recognition for its air purifiers [39,
194 40, 41, 42, 43] as well as for its finest air quality monitor [44]. Data from the biggest network of
195 ground-based sensors is gathered by IQAir. To guarantee data accuracy and dependability, the
196 organization has deployed automated and AI-based tools and sensors (air-quality meter, IOT-
197 based air-quality monitoring system, MQ135 sensors, etc.). Phase8-DHA, Zeenat Block, Dawn
198 Food Head Office, and HAC Agri Limited are the four stations from which data is gathered.

199
200 The hourly concentrations of various pollutants, such as PM2.5, PM10 (particulate matter), NO2
201 (nitrogen dioxide), NH3 (ammonia), and O3 are used to calculate the air quality data (ozone).
202 Along with this, a few additional factors were gathered, including air temperature, air pressure,
203 humidity, wind speed, and wind direction.
204

205 **3.2 Data Pre-processing**
206 An hourly data on a variety of air contaminants was gathered from January 2018 to May 2020.
207 The missing values were substituted by the mean of that column for data variables with missing
208 values of 1000 or less. January 2018 to May 2020 is covered by the total processed data. Table 1
209 lists station details as well as a list of the atmospheric contaminants and the number of missing
210 values for each one. For the purpose of identifying the outlier and imputing its values to the
211 column mean, we employed scatter plots and the Mahalanobis Distance (MD) to discover the
212 outlier.
213

214 **3.3 Machine Learning Techniques**
215 In machine learning, we work with a variety of algorithms that enable us to discover the link
216 between data to produce the final prediction; the type of prediction model where we need to find
217 prediction output in the form of continuous numerical values is known as a regression issue.
218 Modern machine learning techniques, such as CNN, N-BEATS, Decision Tree Regressor,
219 Support Vector Regressor, Random Forest Regressor, Auto-Regressive Integrated Moving-
220 Average (ARIMA), and LSTM, have been used to explore and analyze the changes in
221 atmospheric pollutants that occurred during the COVID-19 lockdown.
222

223 **3.3.1 Decision Tree Regressor**
224 The main idea behind a regressor tree is to discover the point in the independent variable and
225 divide the dataset into two parts so that the mean square error is minimized at that point. The
226 algorithm then repeatedly accomplishes this and produces a tree-like structure.
227

228 **3.3.2 Support Vector Regressor**
229 A supervised learning approach called the support vector regressor (SVR) is used to forecast
230 discrete values. The hyper-plane that maximizes the number of points is the best-fit line, as
231 determined by SVR. similar to other regression methods that aim to reduce the discrepancies
232 between expected and actual data. Within the threshold values, SVR locates the line with the
233 greatest match. The distance between the boundary line and the hyper-plane is known as the
234 threshold value. SVR's fit time-complexity (number of samples) is slightly over the quadratic,
235 which makes it challenging to scale with the dataset (having more than 1 thousand samples).
236

237 **3.3.3 Random Forest Regressor**

238 Regression and classification issues are solved using decision trees. In the case of regression, the
239 data flow is modelled as a tree; it begins at the root node and proceeds through splits based on
240 the outcome variable until it reaches the leaf node, where the findings are shown. Decision trees
241 are further divided into ensemble and bootstrapping. Bootstrapping is a random-forest approach
242 that builds numerous random decision trees from the data by combining ensemble learning with
243 the decision tree. We have an output that can be used for strong classification or regression after
244 averaging.
245

3.3.4 Convolution Neural Network (CNN)

247 The fact that CNN supports dilated convolution, in which the filters are utilized to calculate the
248 dilations between the cells, allows it to handle time series problems and perform tasks linked to
249 time-series forecasting. The neural network can comprehend the association between
250 observations of time-series data due to the magnitude of the space between the cells. The CNN
251 architecture is capable of integrating feature-engineering into a single framework, feature
252 extraction, and time-series representation of the data. CNN is regarded as being able to extract
253 deep characteristics that are independent of time and being very noise resistant.
254

3.3.5 N-BEATS

256 A deep stack of ensemble feed-forward networks connecting forecast and back-cast connections
257 is the basis of the deep learning model N-BEATS. The accurate successive block model takes
258 into account the mistake resulting from the reconstruction of the back-cast from the previous
259 block and changes projections accordingly. Similar to the "Box-Jenkins" approach, this
260 procedure fits the ARIMA model. The inner-learning and outer-learning operations make up the
261 N-BEATS system. The inner-procedure works within the blocks and aids in local-temporal value
262 capture in the model. The outer-learning process takes place within the stacks and aids the model
263 in picking up on global characteristics.
264

3.3.6 ARIMA

266 The ARIMA model uses prior values and its own lag predicting mistakes to describe a particular
267 time series. Three terms, denoted by p, d, and q, make up the ARIMA model.
268 The Autoregressive (AR) term's order is p.
269 The Moving Average (MA) term's order is q.
270 The amount of differencing (d) necessary to make a time series stationary is d.
271 The term "AR" refers to a linear-regression model that calculates its lag values; this model
272 performs best when the predictors are unrelated to one another and do not correlate. If time-
273 series are already stationary, then d = 0, which is the minimal difference necessary for stationary
274 series. The number of delays that were employed to forecast Y is indicated by the auto-regressive
275 (AR) term p. And q is the rank of the moving average (MA), which indicates how many forecast
276 mistakes are included in the ARIMA model.
277

278 **3.3.7 LSTM**

279 The kind of recurrent neural network (RNN) known as long short-term memory (LSTM) is
280 capable of learning order dependency in sequencing prediction tasks. One input layer, one output
281 layer, and a small number of completely linked hidden layers with gate units and matching
282 memory cells make up the LSTM. Each memory cell's internal architecture ensures that the flow
283 of errors inside its constant error-carrousel is constant (CEC). For the purpose of understanding
284 the error-flow inside each memory cell's CEC, two unit-gates open and close. The input
285 multiplicative gate guards against irrelevant data for the CEC. The output multiplicative gate
286 guards against irrelevant memory content for other gates.

287 **Mathematical Model**

288 Let LSTM has input x(t) and the output o(t)

289 h(t-1) and c(t-1) these are inputs from previous time step

290 c(t) and h(t) are consumption of next time-step, as shown in Figure 1.

291 $$f_t = \sigma_g \left(W_f \times x_t + U_f \times h_{t-1} + b_f \right) \qquad (1)$$

292 $$i_t = \sigma_g \left(W_i \times x_t + U_i \times h_{t-1} + b_i \right) \qquad (2)$$

293 $$o_t = \sigma_g \left(W_o \times x_t + U_o \times h_{t-1} + b_o \right) \qquad (3)$$

294 $$c^{\wedge}_t = \sigma_c \left(W_c \times x_t + U_c \times h_{t-1} + b_c \right) \qquad (4)$$

295 $$c_t = f_t \cdot c_{t-1} + i_t \cdot c^{\wedge}_t) \qquad (5)$$

296 $$h_t = o_t \cdot \sigma_c (c_t) \qquad (6)$$

297 Where

298 f_t is the forget gate

299 i_t is the input gate

300 o_t is the output gate

301 c_t is the cell state

302 h_t is the hidden state

303 σ_g is the sigmoid

304 σ_c is the tanh

305 . is the element wise multiplication

306

307 LSTM equations also generated f(t), i(t) and c^ these are the internal consumption of LSTM and
308 being used for generating c(t) and h(t).

309 The above mathematical model shows the identical steps for one time-step, meaning all the
310 equations must be recomputed for next time-step. Suppose we have 10 time-steps, we need to
311 recompute the above equation for ten times. (Wf, Wi, Wo, Wc, Uf, Ui, Uo, Uc) all these are
312 weight matrices and (bf, bi, bo, bc) are the biases and these are not time dependent.

313

314 **3.4 Exploratory Analysis of Data**

315 The LSTM model was trained using data from January 2018 to December 2019. But a different
316 collection of special data from January to February 2020, known as the validation set, was
317 utilized for validation (VS). And the Lock-down Set was the data from March 2020 to May

318 2020. (LD). There are 3647 instances in the validation and lock-down collection, which represent
319 the hourly AQI values of various air contaminants. To estimate the values for VS and LD, the
320 already trained model was employed.
321

322 In order to forecast the concentration of particular contaminants, the LSTM model was used.
323 Target and predictive variables serve as the foundation for machine learning and other deep
324 learning approaches [14]. The concentration of certain air pollutants such PM2.5, PM10
325 (particulate matter), NO2 (nitrogen dioxide), and O3 were the goal variables (y) (ozone). These
326 predictive factors helped the LSTM model capture the seasonal trend of traffic flow and
327 industrial operations. The predictive variables (X) were data linked to environmental conditions
328 coupled with their temporal variables. The RMSE, MAE, and R-SQUARE values were used to
329 assess the model's performance.
330

331 The classification of datasets into training, validation, and lockdown time frames is shown in
332 Figure 2 along with dataset details. The LSTM model was trained using time series data from
333 January 2018 to December 2019, the validation set was used for cross-validation from January
334 2020 to February 28, 2020, and the actual values during the lockdown period (March–May) were
335 compared with the model predictions using least square errors. The entire research methodology
336 is shown in Figure 3. First, the dataset is preprocessed, then thoroughly refined, and preprocessed
337 data is passed to the most cutting-edge machine learning models. Next, each of the machine
338 learning models is evaluated based on RMSE, MAE, and R-SQUARE score, and finally the final
339 results are compared. The model with the lowest error score is considered to be the best model.
340

341 **3.5 Inferential statistical analysis**
342 To infer attributes from the dataset, inferential statistical analysis is employed. These
343 characteristics aid in predicting future values. Different hypotheses can be tested; inferential
344 statistics heavily relies on the testing of hypotheses. It enables us to judge if the pattern we notice
345 is the result of chance or whether it actually has statistical significance. In order to achieve this,
346 we must first create a hypothesis and a null hypothesis. The hypothesis is the trend we expect to
347 see in the data, while the null hypothesis is the polar opposite. In our investigation, we focused
348 on four air pollutants: PM2.5, PM10, NO2, and O3. We'll analyze how these contaminants
349 impact the air quality index (AQI). According to our theory, the concentration of these
350 contaminants and the AQI are linearly correlated. The null hypothesis demonstrates that there is
351 no linear relationship between these pollutants' concentrations and AQI. Following that, we do
352 various tests to determine if there is statistical significance. Table 2 demonstrates that the p-value
353 for our parameters is less than 0.05 and that the F-statistics for PM2.5 is substantial. The null
354 hypothesis may thus be rejected as we can presume that there is statistical significance. Similar
355 circumstances apply to the pollutants PM10, NO2, and O3. In Table 2, the negative slope
356 indicates that the air quality index (AQI) will decline as pollutant concentration rises and
357 increase as concentration falls.
358

359

3.6 Validation method and comparison

360 The projected concentration of each air contaminant is shown by the LSTM findings. These
361 "predicted concentrations" and the contaminants' "actual values" were contrasted. Both from a
362 long-term and short-term viewpoint, the evolution of the historical data was compared to our
363 technique. Due to a decrease in traffic and industrial activity, we see a decrease in the
364 concentration of air pollutants.

366

4 Experiment and Results

368 The key concern is using deep learning to research how COVID-19 lockdown affects air quality.
369 The algorithms use historical data from January 2018 to May 2020 to forecast the air quality
370 during the shutdown. The historical records were further divided into two categories: lockdown
371 set, and validation set. The historical records prior to March 2020 were used for validation, and
372 the data from March through May 2020 displays the lockdown set that is used for testing. The
373 data from January 2018 through December 2019 was used for training the models. Models are
374 employed to compare each pollutant's projected levels to its actual values. The RMSE, MAE,
375 and R-SQUARE values are used to assess models. In comparison to other models, the LSTM
376 model demonstrated the fewest errors and the best predictions for the lockdown and validation
377 time frames.

378

4.1 Explorative Analysis

380 We use explorative data analysis to determine the relative importance of each air contaminant in
381 order to comprehend the relationships between them. We employ principal component analysis
382 for that (PCA). For the best outcomes, it is advised to select the components that account for
383 between 70 and 80 percent of the overall variation based on "explained variance". We select five
384 elements from the data that are the most beneficial and also include important information by
385 keeping the concept in mind.

386

4.2 Hyper-parameter tuning for LSTM

388 To influence the behavior of deep learning and machine learning models, hyper-parameter
389 tweaking is crucial. Our model's ability to produce accurate results depends on our ability to
390 fine-tune its hyper-parameters. If we are unable to minimize the loss-function, our model will
391 produce more mistakes. We have adjusted the hyper-parameters for the LSTM model while
392 keeping this idea in mind. The LSTM model has four layers: one input layer, two hidden layers,
393 and one output layer. It is generally recommended to utilize two hidden layers for complicated
394 problems, and we have done the same. Second, we concentrate on the number of nodes. When
395 we lower the number of nodes, it improves accuracy and, in other words, reduces overfitting.
396 Removing nodes act as a regularizer. 50 nodes were employed in the input layer, while 50%
397 fewer nodes were used in the second and third layers. We used the ReLU activation function on
398 the hidden layers because ReLU is nonlinear and has the advantage of not having back

399 propagation error, and we used the tanh activation function on the output layer because the
400 output from that layer ranges from -1 to 1, and tanh is thought to be a good one for solving
401 regression problems. In our situation, after 100 epochs the validation accuracy started to fall,
402 therefore, we set epochs to 100. As Adam optimizer incorporates the best qualities of RMSProp
403 and AdaGrad algorithms, it can handle the sparse gradients on noisy data extremely quickly. The
404 learning rate was set to 0.1. We utilized root mean square error and the Adam optimizer as
405 measurements. Although we have used LSTM with a variety of settings, the results are greatest
406 when using the fine-tuned parameters stated above.
407
408 **4.3 Evaluation Parameters**
409 We employed several machine learning assessment criteria. RMSE, MAE, and R-SQUARE
410 score were among the metrics used.
411
412 **4.3.1 Root Mean Square Error**
413 The standard-deviation of the prediction errors is known as RMSE (known as residuals). The
414 prediction errors show how far apart from the regression line the data points are. The
415 concentration of data around the best-fit line is measured by RMSE.
416
417 $$RMSE\ f_o = \sqrt[2]{[\sum_{i=1}^{N}(Z_f i - Z_o i)^2 / N]} \qquad (7)$$
418
419 Where: $\sum$ = summation ("add up")
420 $(Z\ fi – Zoi)^2$ = differences, squared
421 N = Normal size
422
423 **4.3.2 Mean Absolute Error**
424 Our second evaluation parameter is mean absolute error, and absolute error is defined as the
425 "magnitude of difference between the mean value and each individual value." It can also be
426 defined as "the arithmetic mean of all the absolute errors is taken as the mean absolute error of
427 the value of that quantity."
428
429 $$MAE = 1/n\sum_{i=1}^{n}|x_i - x| \qquad (8)$$
430
431 Where:
432 $|x_i – x|$ = absolute-error
433 n = number of errors
434
435 **4.4 R-SQUARE Score**
436 The third parameter is the R-SQUARE score, defined as "the statistical measures that represent
437 the proportion of the variance for a dependent variable that is explained by an independent
438 variable or a variable in a regression model." In contrast, the correlation shows a strong

439   relationship between the dependent and independent variables. It also shows the extent to which
440   a variable explains the variance of another variable.

441

442    $R^2 = SS_{regression} / SS_{total}$ (9)

443

444   Where
445   SS $_{regression}$ is the sum squares by the regression
446   SS $_{total}$ is the total sum of squares

447

448   **4.5 Deep Learning Results**
449   We have trained various machine learning models, including Decision Tree, SVR, Random
450   Forest, ARMIA, CNN, N-BEATS, and LSTM, on the entire training dataset, which contains
451   historical records from January 2018 to December 2019, in order to better understand the true
452   changes in the concentration of various atmospheric pollutants. Additionally, these models were
453   fitted independently to the data from January 2020 to May 2020 (validation set and lockdown
454   set). The results were then assessed using the Mean Absolute Error (MAE), R-SQUARE Score,
455   and Root Mean Square Error (RMSE). When compared to other models, the LSTM model
456   performs best and has the fewest mistakes. The primary explanation for this is because LSTM
457   excels in time-series forecasting and develops forecasts based on prior sequential data.
458   Compared to other models, LSTM has a remarkable capacity to efficiently remember data.

459

460   Figure 4 displays the LSTM model's training and validation losses in terms of Mean Square
461   Error and Root Mean Square Error. On one hundred epochs, the initial training was run. After 10
462   epochs, the training and validation losses are consistent. During the first stages of training, the
463   error was at its highest point and the gap between the two losses was at its greatest.

464

465   **4.6 Comparison of LSTM model with other Machine learning models**
466   In relation to various pollutants, the overall findings of several machine learning models are
467   shown in Table 3. On the air quality dataset, we applied a variety of machine learning models,
468   including Decision Tree, SVR, Random Forest, ARIMA, CNN, N-BEATS, and LSTM. Error
469   metrics including RMSE, MAE, and R-SQUARE score were used to evaluate the machine
470   learning models. The top four pollutants accounted for 70% of the overall variance when we
471   used PCA to choose features. The contaminants that were chosen were PM2.5, PM10, NO2, and
472   O3. For each pollutant, we have utilized a scatter plot and the Mahalanobis Distance (MD) to
473   identify the outlier and impute its values using the meaning of the column. For each pollutant, we
474   modified the LSTM model's parameters and put other cutting-edge machine learning models into
475   practice. The findings were then created based on error measurements. All of the machine
476   learning models worked well, however the LSTM model excelled in comparison to the others.
477   The LSTM model has the lowest error for each of the various pollutants when compared to other
478   models. Since we are projecting the pollutant concentration and utilizing time series data, the

479 LSTM network is the ideal model for time series data. LSTM has the capacity to predict future
480 values based on prior sequential data (past historical records). When compared to other machine
481 learning models, LSTM performed better since it can anticipate the future while learning the
482 prior levels of various contaminants. PM2.5 pollutant had a score of 4.46 percent MAE, 0.25
483 percent R-SQUARE score, and 9.99 percent RMSE from the LSTM model. The PM10
484 pollutant's LSTM score was 12.84 percent RMSE, 0.75 percent R-SQUARE score, and 8.58
485 percent MAE. The results of the LSTM model for NO2 were 4.35 MAE, 0.12 R-SQUARE, and
486 6.77 RMSE. 8.20 percent MAE, 0.17 percent R-SQUARE, and 29.79 percent RMSE made up
487 the O3 LSTM score.

488

489 The four parts that make up an elementary LSTM unit are as follows: a cell, an input gate, an
490 output gate, and a forget gate. A cell's function is to retain values for any given period of time.
491 Additionally, gates' function is to control how information enters and exits a cell. For time-series
492 forecasting, LSTM consistently delivered good results. Sequential data's future values are
493 effectively predicted. There is no need for precise modification because the LSTM gives us a
494 wealth of parameters, including input, output, biases, and learning rates. The LSTM model
495 outperforms ARMIA in terms of performance. The major causes of this are because ARMIA is
496 computationally costly and performs poorly for situations involving seasonal time-series and
497 long-term forecasting. While LSTM often works with large datasets, where sufficient data must
498 be provided for initial model training, ARIMA excels on smaller datasets, but it might be
499 challenging to pinpoint the ARIMA model's pivotal moments. When it comes to decision trees
500 and random forests, a forest needs a lot of memory since it stores data from hundreds of trees.
501 The LSTM cell stores information, and gates control that memory. Similarly, visualize the series
502 of choices that results in computational burden. That is a reliable answer. The SVM method
503 works well for small datasets, but it struggles with noisy data, data where target classes overlap,
504 or data where the number of training data samples is insufficient for the number of features. As a
505 result, these algorithms do not perform as well as LSTM for air quality datasets.

506

507 **4.7 LSTM Results with respect to different pollutants**
508 Figure 5 shows the PM2.5 concentration for the period from January to May. The predicted
509 values, which were compared to the actual values. The time series data from January to May are
510 displayed on the plot. The validation set is from January 2020 to February 28th, and the
511 lockdown period is from March 2020 to May 2020. The graph's analysis reveals that there is no
512 average decrease in PM2.5 concentration from January to the end of February. However, there is
513 a decrease in the concentration of PM2.5 after the middle of March (the lockdown timeframe).

514

515 The concentration of NO2 for the period from January to May is shown in Figure 6. The
516 predicted values, assessed against the actual values. The validation set is from January through
517 February 28th, and the lockdown period is from March through May of the following year. The
518 concentration of NO2 decreases from the middle of March to the end of May. The graph reveals

519 a steady decline in air pollution after March, which appears to be the true result of the COVID-
520 19 lockout.
521
522 Figure 7 displays the PM10 concentration. The predicted values, which were compared to the
523 actual values. The figure displays the time series data as well as the PM10 concentration. The
524 validation set is the period from January 2020 to February 2020; the lockdown period is from
525 March 2020 to May 2020. It was noticed that the amount of PM10 pollutants was falling.
526
527 Figure 8 displays the Ozone (O3) levels for various time periods. The predicted values, which
528 were compared to the actual values. The statistics for several time periods from January through
529 May are represented on the figure. The validation set is from January through February 28th, and
530 the lockdown period is from March through May of the following year. The graphic
531 unequivocally demonstrates that the amount of ozone is rising (O3). The quality of ozone has
532 significantly improved as a result of the concentration of some air contaminants being lower.
533 There is a decrease in the concentration of air contaminants, as seen in Figures 3, 6, and 7.
534

4.8 Forecasting Moving Average

536 The first stage in the forecasting process is to ascertain the maximum number of hours the model
537 can anticipate accurately. Most of the time, the extended lead-time may cause performance that
538 is ineffective and inaccurate. The R-SQUARE error levels determine how accurate the results
539 are. R-SQUARE error levels increase as predicting time increases. On the other side, a shorter
540 predicting period results in the least number of mistakes. In light of this, we have predicted the
541 concentration of several pollutants, including PM2.5, PM10, NO2, and O3, over the next four
542 days using a moving average.
543
544 The 4-day moving average of PM2.5 and PM10 pollutants is shown in Figures 9(a) and 9(b), and
545 the findings show a reduction in PM2.5 and PM10 pollutant concentrations for the next four
546 days. The error measurements have not been impacted by the moving average. The LSTM model
547 still shows minimum errors in terms of RMSE, MAE, and R-SQUARE score while forecasting 4
548 days of moving average.
549
550 Figure 10 (a) depicts a drop in NO2 concentration and Figure 10 (b) depicts an increase in O3
551 concentration (ozone). In terms of RMSE, MAE, and R-SQUARE score, the 4-day prediction of
552 moving average produces strong results with little mistakes.
553

4.9 Reduction in Air Pollution

555 Comparing the lockdown of 2020 to the same time period in the preceding years, Table 4
556 provides a clearer picture of certain contaminants and their average concentrations.
557 We calculated the average concentration for each pollutant and displayed it in Table 4. For our
558 research investigation, the well-known contaminants that were thought to be the more serious

559 causes of pollution were selected. The contaminants that were considered were PM10, PM2.5,
560 NO2, NOx, and O3. These contaminants' typical concentrations were measured and contrasted
561 with values from the same period in the prior years. Table 4 displays time series data for three
562 months and compares the average concentration of various pollutants during the lockdown
563 periods (March 2020 to May 2020) to the corresponding periods for the two years prior, 2018
564 and 2019, as well as the average concentration of various pollutants chosen for the years of 2018,
565 2019, and 2020. (March, May, and April). We found that the concentration of PM2.5 pollutants
566 was 31% lower in March 2020 than it was in March 2019 after examining these statistical data.
567
568 When the PM2.5 pollutant concentrations in April (2019 and 2020) were compared, April 2020
569 had a 33 percent lower pollutant concentration than April 2019. When May 2019 and May
570 2020's PM2.5 concentrations were compared, it was found that the latter month's reading was 57
571 percent lower. Similar to March 2019, there was a 7.8% decrease in PM10 concentration in
572 March 2020 compared to March 2019. When compared to April 2019, there was a 42 percent
573 decrease in April 2020. Comparing March 2020 to March 2019, the concentration of NO2
574 declined to 1.92 percent, by 43 percent in April 2020, and by 67 percent in May 2020. In
575 comparison to March 2019, April 2020, and May 2019, the average NOx concentration reduced
576 by 11.1 percent, 40 percent, and 70 percent, respectively. The quality of the air will improve as
577 the number of atmospheric pollutants decreases, improving the overall concentration of ozone
578 O3 during the lockdown period. In comparison to March 2019, there was a 38 percent rise in
579 ozone (O3) in March 2020. Ozone (O3) levels increased for the month of April 2020 by 9.3%
580 compared to April 2019. In addition, ozone (O3) increased by 12 percent in May 2020 compared
581 to May 2019 within the same calendar year. When we compared the levels of several
582 contaminants in 2020 and 2018, there was a considerable drop. Figure 11 provides a clear visual
583 illustration of Table 4's graphical form.
584
585 Figure 11 shows a visual depiction of the average concentration of several air pollutants during
586 the months of March, April, and May for the years (2018), 2019 and 2020.
587

5 Conclusion and Future Work

589 In this work, we examined variations in the concentration of several air pollutants during the
590 Lahore COVID-19 lockdown (Pakistan). The LSTM model was used in the study to analyze the
591 discrepancy between real and forecasted values for the concentration of various contaminants.
592 The LSTM model's prediction was more than enough for examining variations in the
593 concentration of several contaminants. During the lockdown, the number of pollutants was
594 significantly reduced, while the air's quality (Ozone O3) was enhanced. Nevertheless, because of
595 decreased industrial activity and traffic during the lockdown in the Pakistani city of Lahore, the
596 "Air Quality Index" as a whole improved. The findings show that the LSTM model has high
597 potential and is regarded as a suitable tool for investigating and assessing the changes occurring
598 to air contaminants during the COVID-19 shutdown. We came to the conclusion that deep

599 learning is the most reliable approach for researching and monitoring time series issues.
600 However, further research is needed to examine the changes taking place in time series issues on
601 a larger scale. We will research deep learning models with enhanced generalization capabilities
602 in the future to address time-series issues. Moreover, we intend to make a generic prediction of
603 the air quality by taking the dataset of post covid from different cities of Pakistan.

604

605 References

606 [1] N. Cobb, A. Papali, L. Pisani, M. J. Schultz, and J. C. Ferreira, "Pragmatic recommendations
607 for infection prevention and control practices for healthcare facilities in low- And middle-income
608 countries during the COVID-19 pandemic," American Journal of Tropical Medicine and
609 Hygiene, vol. 104, no. 3, pp. 25–33, 2021.
610 [2] J. Aljohani, J. Shuja, W. Alasmary, and A. Alashaikh, "Evaluating the dynamics of bluetooth
611 low energy based covid-19 risk estimation for educational institutes," Sensors, vol. 21, no. 19,
612 2021.
613 [3] E. Conticini, B. Frediani, and D. Caro, "Can atmospheric pollution be considered a co-factor
614 in extremely high level of SARS-CoV-2 lethality in Northern Italy?," Environmental Pollution,
615 vol. 261. 2020.
616 [4] O. Y. Bello-Chavolla et al., "Unequal Impact of Structural Health Determinants and
617 Comorbidity on COVID-19 Severity and Lethality in Older Mexican Adults: Considerations
618 beyond Chronological Aging," The Journals of Gerontology: Series A, vol. 76, no. 3, pp. E52–
619 E59, 2021.
620 [5] Y. Zhu, J. Xie, F. Huang, and L. Cao, "Association between short-term exposure to air
621 pollution and COVID-19 infection: Evidence from China," Science of The Total Environment,
622 vol. 727, 2020.
623 [6] A. Srivastava, "COVID-19 and air pollution and meteorology-an intricate relationship: A
624 review," Chemosphere, vol. 263. 2021.
625 [7] U. A. Bhatti, Z. Zeeshan, M. M. Nizamani, S. Bazai, Z. Yu, and L. Yuan, "Assessing the
626 change of ambient air quality patterns in Jiangsu Province of China pre-to post-COVID-19,"
627 Chemosphere, vol. 288, 2022.
628 [8] J. Zhang, H. Li, L. Chen, X. Huang, W. Zhang, and R. Zhao, "Particle composition, sources
629 and evolution during the COVID-19 lockdown period in Chengdu, southwest China: Insights
630 from single particle aerosol mass spectrometer data," Atmospheric Environment, vol. 268, 2022.
631 [9] D. Rodríguez-Urrego and L. Rodríguez-Urrego, "Air quality during the COVID-19: PM2.5
632 analysis in the 50 most polluted capital cities in the world," Environmental Pollution, vol. 266.
633 2020.
634 [10] A. Kerimray et al., "Assessing air quality changes in large cities during COVID-19
635 lockdowns: The impacts of traffic-free urban conditions in Almaty, Kazakhstan," Science of The
636 Total Environment, vol. 730, 2020.

637 [11] S. E. Sbai, N. Mejjad, A. Norelyaqine, and F. Bentayeb, "Air quality change during the
638 COVID-19 pandemic lockdown over the Auvergne-Rhône-Alpes region, France," Air Quality,
639 Atmosphere and Health, vol. 14, no. 5, pp. 617–628, 2021.

640 [12] M. Mahmud and E. Riley, "Household response to an extreme shock: Evidence on the
641 immediate impact of the Covid-19 lockdown on economic outcomes and well-being in rural
642 Uganda," World Development, vol. 140, 2021.

643 [13] A. Desvars-Larrive et al., "A structured open dataset of government interventions in
644 response to COVID-19," Scientific Data, vol. 7, no. 1, 2020.

645 [14] I. Šimić, M. Lovrić, R. Godec, M. Kröll, and I. Bešlić, "Applying machine learning methods
646 to better understand, model and estimate mass concentrations of traffic-related pollutants at a
647 typical street canyon," Environmental Pollution, vol. 263, 2020.

648 [15] N. Zhao, G. Wang, G. Li, J. Lang, and H. Zhang, "Air pollution episodes during the
649 COVID-19 outbreak in the Beijing–Tianjin–Hebei region of China: An insight into the transport
650 pathways and source distribution," Environmental Pollution, vol. 267, 2020.

651 [16] F. Liu et al., "Abrupt decline in tropospheric nitrogen dioxide over China after the outbreak
652 of COVID-19," Science Advances, vol. 6, no. 28, 2020.

653 [17] Z. S. Venter, K. Aunan, S. Chowdhury, and J. Lelieveld, "COVID-19 lockdowns cause
654 global air pollution declines," Proceedings of the National Academy of Sciences U. S. A., vol.
655 117, no. 32, pp. 18984–18990, 2020.

656 [18] M. A. A. Al-qaness, H. Fan, A. A. Ewees, D. Yousri, and M. Abd Elaziz, "Improved ANFIS
657 model for forecasting Wuhan City Air Quality and analysis COVID-19 lockdown impacts on air
658 quality," Environmental Research, vol. 194, 2021.

659 [19] W. Mao, W. Wang, L. Jiao, S. Zhao, and A. Liu, "Modeling air quality prediction using a
660 deep learning approach: Method optimization and evaluation," Sustainable Cities and Society,
661 vol. 65, 2021, doi: 10.1016/j.scs.2020.102567.

662 [20] K. Zhang and S. Batterman, "Near-road air pollutant concentrations of CO and PM2.5: A
663 comparison of MOBILE6.2/CALINE4 and generalized additive models," Atmospheric
664 Environment, vol. 44, no. 14, pp. 1740–1748, 2010.

665 [21] M. Bauwens et al., "Impact of Coronavirus Outbreak on NO2 Pollution Assessed Using
666 TROPOMI and OMI Observations," Geophysical Research Letters, vol. 47, no. 11, 2020.

667 [22] T. Le et al., "Unexpected air pollution with marked emission reductions during the COVID-
668 19 outbreak in China," Science, vol. 369, no. 6504, pp. 702–706, 2020.

669 [23] R. V. Gagliardi and C. Andenna, "A machine learning approach to investigate the surface
670 ozone behavior," Atmosphere (Basel)., vol. 11, no. 11, 2020.

671 [24] P. Krecl, A. C. Targino, G. Y. Oukawa, and R. P. Cassino Junior, "Drop in urban air
672 pollution from COVID-19 pandemic: Policy implications for the megacity of São Paulo,"
673 Environmental Pollution, vol. 265. 2020.

674 [25] G. Dantas, B. Siciliano, B. B. França, C. M. da Silva, and G. Arbilla, "The impact of
675 COVID-19 partial lockdown on the air quality of the city of Rio de Janeiro, Brazil," Science of
676 The Total Environment, vol. 729, 2020.

677 [26] L. Li et al., "Air quality changes during the COVID-19 lockdown over the Yangtze River
678 Delta Region: An insight into the impact of human activity pattern changes on air pollution
679 variation," Science of The Total Environment, vol. 732, 2020.

680 [27] P. Wang, K. Chen, S. Zhu, P. Wang, and H. Zhang, "Severe air pollution events not avoided
681 by reduced anthropogenic activities during COVID-19 outbreak," Resources, Conservation and
682 Recycling., vol. 158, 2020.

683 [28] R. Tosepu et al., "Correlation between weather and Covid-19 pandemic in Jakarta,
684 Indonesia," Science of The Total Environment, vol. 725, 2020.

685 [29] Y. Ma et al., "Effects of temperature variation and humidity on the death of COVID-19 in
686 Wuhan, China," Science of The Total Environment, vol. 724, 2020.

687 [30] D. Cao, S. Ji, R. Cui, and Q. Liu, "Multi-task learning for digital rock segmentation and
688 characteristic parameters computation," Journal of Petroleum Science and Engineering, vol. 208,
689 2022.

690 [31] W. Wang, Z. Xu, and J. W. Lu, "Three improved neural network models for air quality
691 forecasting," Engineering Computations, vol. 20, no. 1–2, pp. 192–210, 2003.

692 [32] G. Corani, "Air quality prediction in Milan: Feed-forward neural networks, pruned neural
693 networks and lazy learning," Ecological Modelling, vol. 185, no. 2–4, pp. 513–529, 2005.

694 [33] Y. S. Chang, H. T. Chiao, S. Abimannan, Y. P. Huang, Y. T. Tsai, and K. M. Lin, "An
695 LSTM-based aggregated model for air pollution forecasting," Atmospheric Pollution Research,
696 vol. 11, no. 8, pp. 1451–1463, 2020.

697 [34] H. Karimian, Q. Li, C. Li, L. Jin, J. Fan, and Y. Li, "An improved method for monitoring
698 fine particulate matter mass concentrations via satellite remote sensing," Aerosol and Air Quality
699 Research, vol. 16, no. 4, pp. 1081–1092, 2016.

700 [35] X. Shi and G. P. Brasseur, "The Response in Air Quality to the Reduction of Chinese
701 Economic Activities During the COVID-19 Outbreak," Geophysical Research Letters, vol. 47,
702 no. 11, 2020.

703 [36] S. Mahato, S. Pal, and K. G. Ghosh, "Effect of lockdown amid COVID-19 pandemic on air
704 quality of the megacity Delhi, India," Science of The Total Environment., vol. 730, 2020.

705 [37] L. Menut, B. Bessagnet, G. Siour, S. Mailler, R. Pennel, and A. Cholakian, "Impact of
706 lockdown measures to combat Covid-19 on air quality over western Europe," Science of The
707 Total Environment, vol. 741, 2020.

708 [38] Hassan, M., Chen, J., Zhu, C., & Zukaib, U. (2022, May). Adoption of Blockchain-based
709 Artificial Intelligence in Healthcare. In 2022 5th International Conference on Artificial
710 Intelligence and Big Data (ICAIBD) (pp. 140-144). IEEE.

711 [39] Asarch, S., Glendon, H., & @tomsguide. (2022, August 24). The best air purifiers in 2022
712 tested and rated | Tom's Guide. The Best Air Purifiers in 2022 Tested and Rated;
713 www.tomsguide.com. https://www.tomsguide.com/best-picks/best-air-purifiers

714 [40] Parent Tested Parent Approved empowers consumers to make informed purchasing
715 decisions. (2020, July 30). PTPA Media; www.ptpaawards.com. https://www.ptpaawards.com/

716   [41] 53 Companies Honored with gia Awards for Product Design - International Housewares
717   Association. (2019, March 3). International Housewares Association; blog.housewares.org.
718   https://blog.housewares.org/2019/03/02/13-companies-honored-gia-awards-product-design/
719   [42] Morgan, G. (2019, August 19). Announcing The Winners Of The 2019 TWICE VIP
720   Awards! TWICE; www.twice.com. https://www.twice.com/vip-awards/announcing-the-winners-
721   of-the-2019-twice-vip-awards
722   [43] Parent Approved, P. T. (2020, September 22). Atem Personal Air Purifier | PTPA | Parent
723   Tested Parent Approved. Atem Personal Air Purifier | PTPA | Parent Tested Parent Approved;
724   www.ptpa.com. https://www.ptpa.com/product/atem-personal-air-purifier
725   [44] BK E-Magazine 781, Feb 28 2020. (n.d.). BK E-Magazine 781, Feb 28 2020 by BK
726   Magazine - Issuu; issuu.com. Retrieved August 25, 2022, from
727   https://issuu.com/bkmagazine/docs/pdf-781-feb-28-20.

Figure 1

LSTM Mathematical Form

Mathematical Model

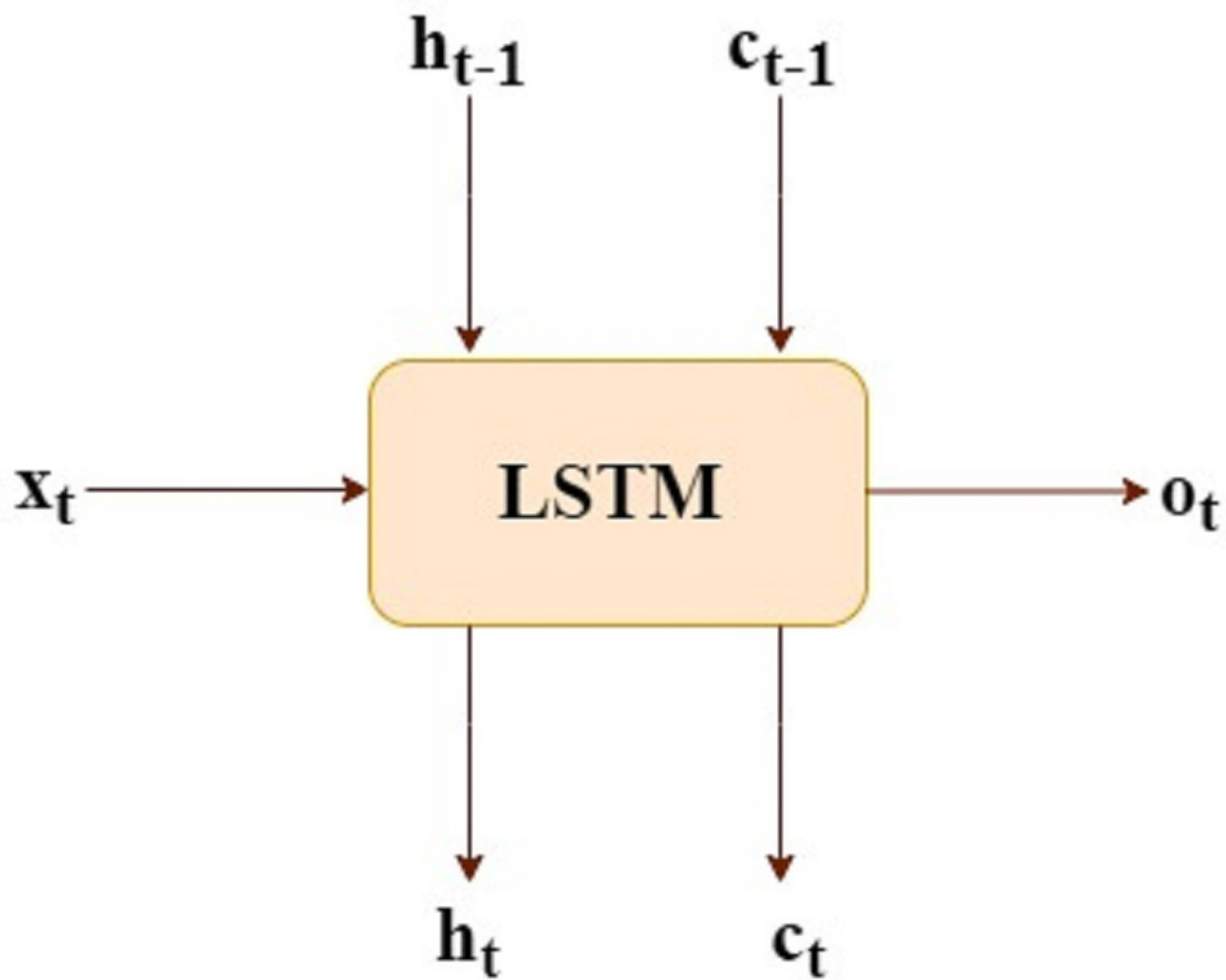



Figure 2

Dataset information and categorization of dataset

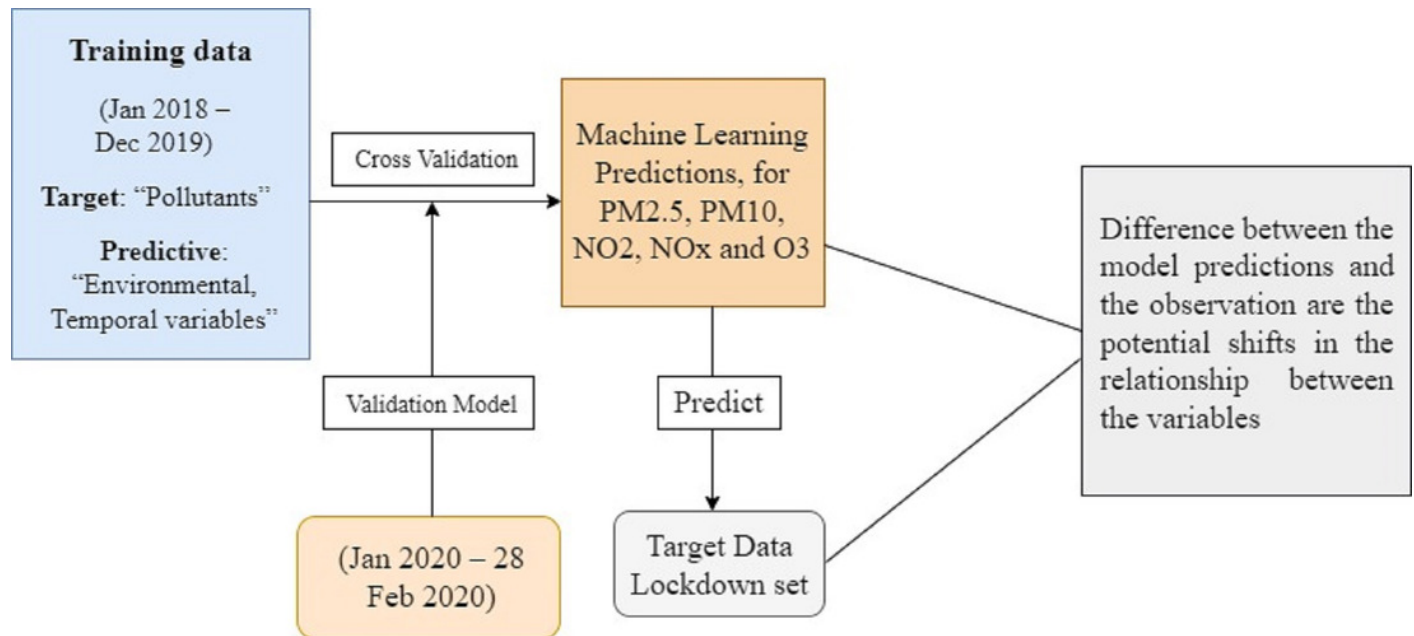



Figure 3

Research methodology for air quality analysis using machine learning techniques

Research Methodology for air quality analysis using machine learning techniques

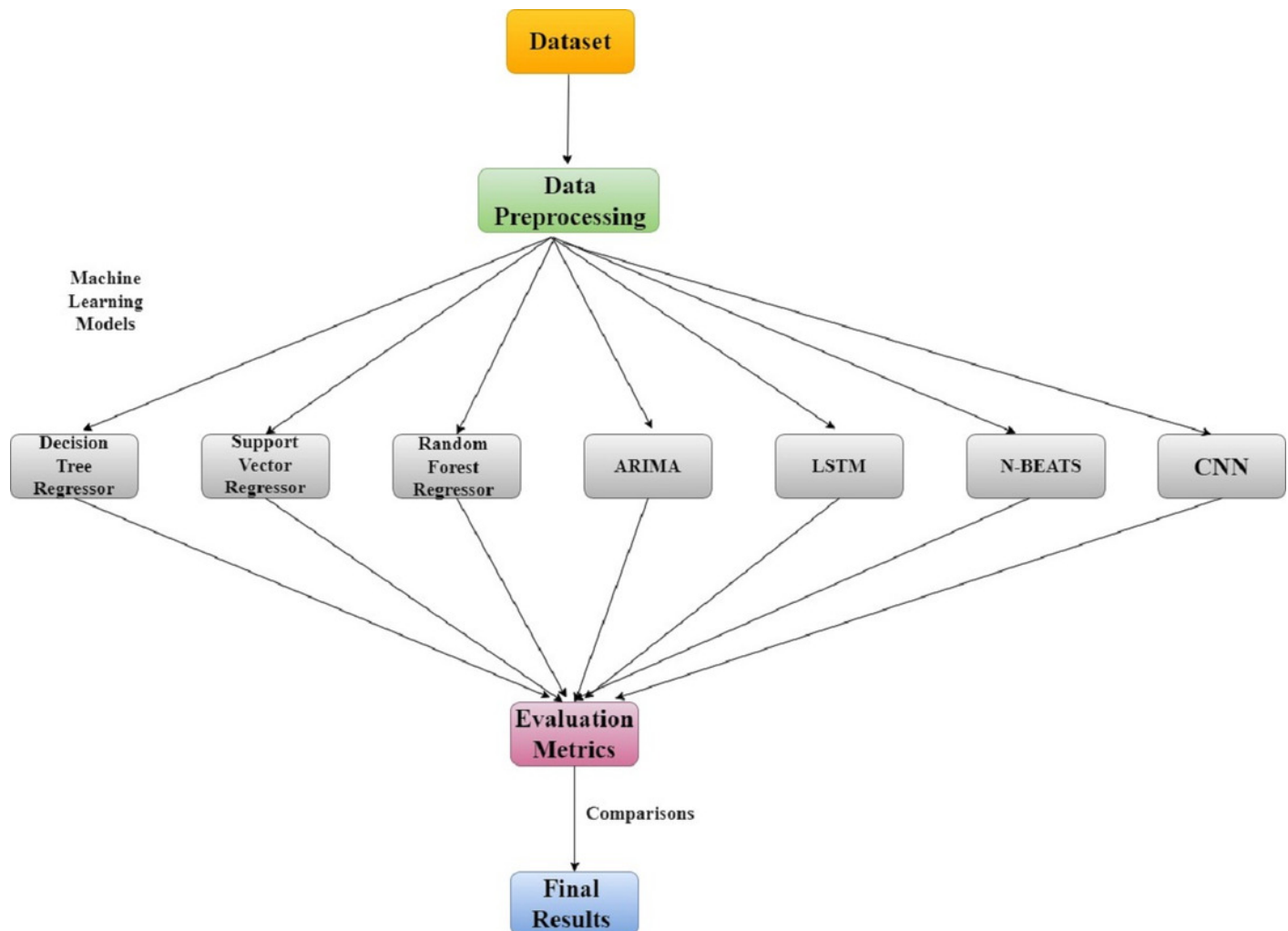



Figure 4

LSTM training and error minimization against 100 epochs

Figure 4. LSTM training and error minimization against 100 epochs

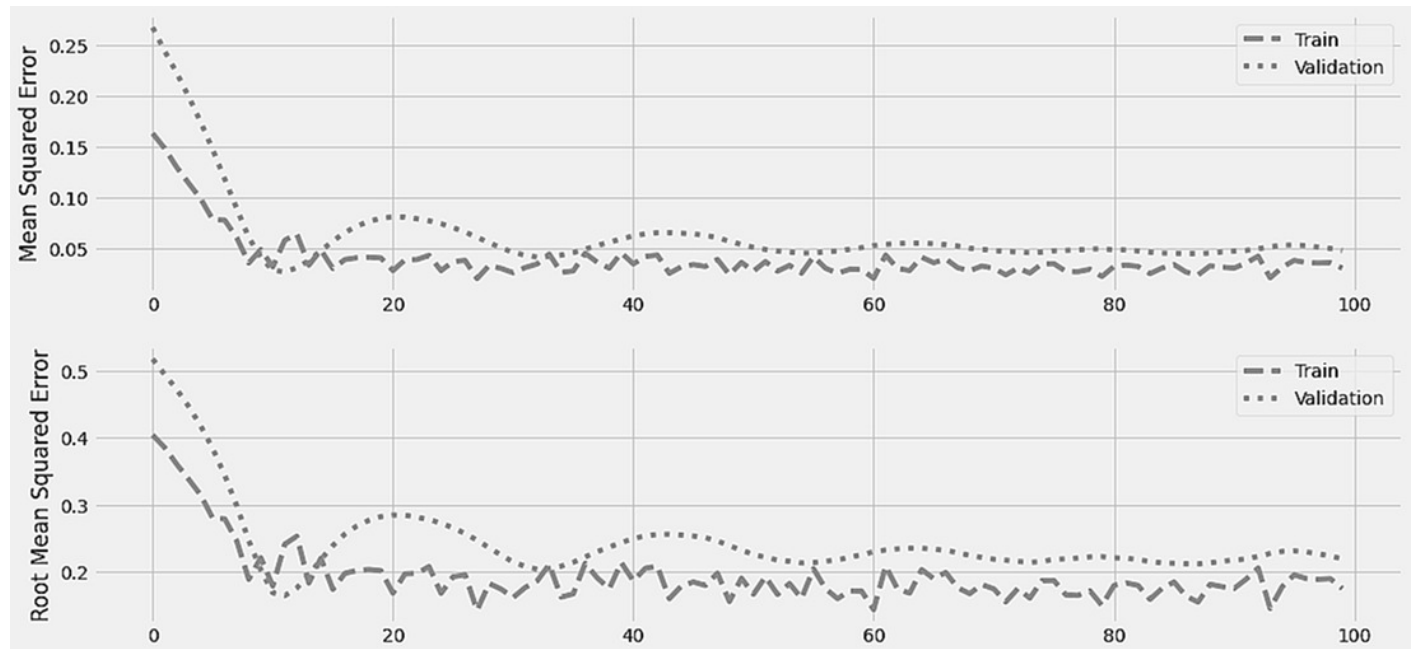



Figure 5

Time Series plot for the concentration of PM2.5 pollutant (Jan 2020 – May 2020)

Time Series plot for the concentration of PM2.5 pollutant

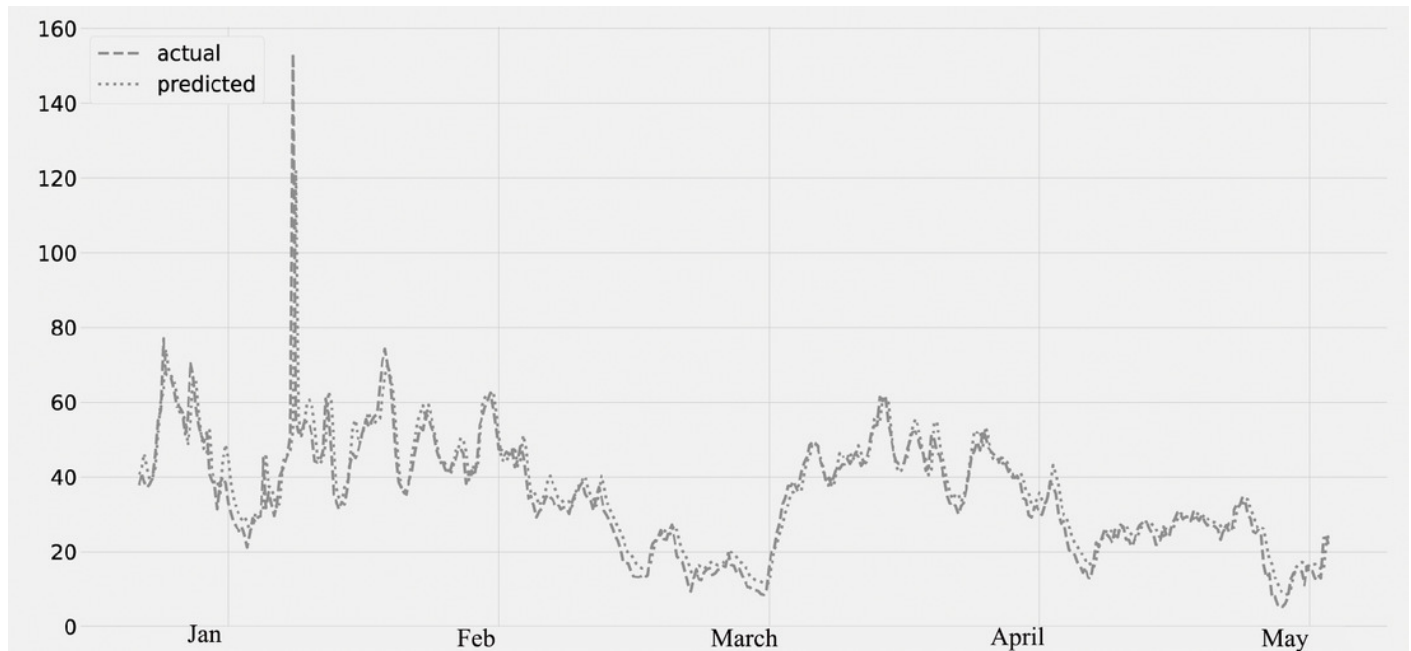



Figure 6

Time Series plot for the concentration of NO2 pollutant (Jan 2020 – May 2020)

Time Series plot for the concentration of NO2 pollutant

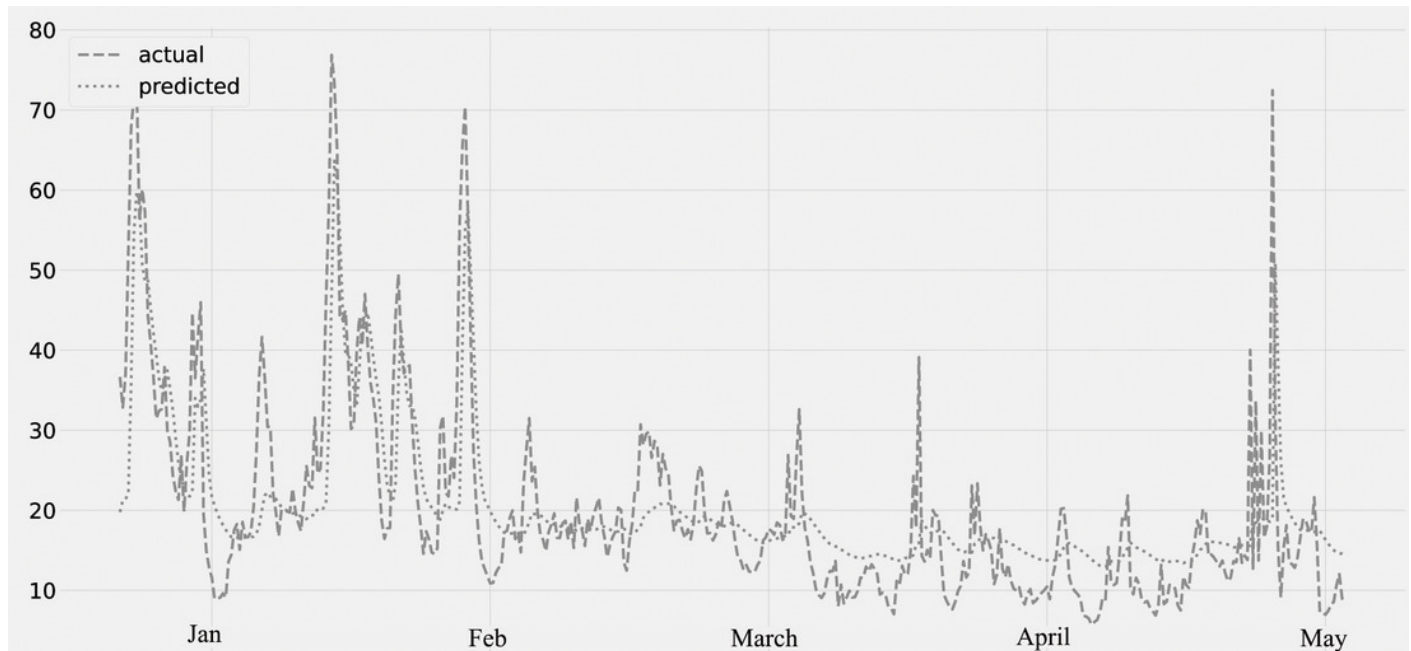



Figure 7

Time Series plot for the concentration of PM10 pollutant (Jan 2020 – May 2020)

Time Series plot for the concentration of PM10 pollutant

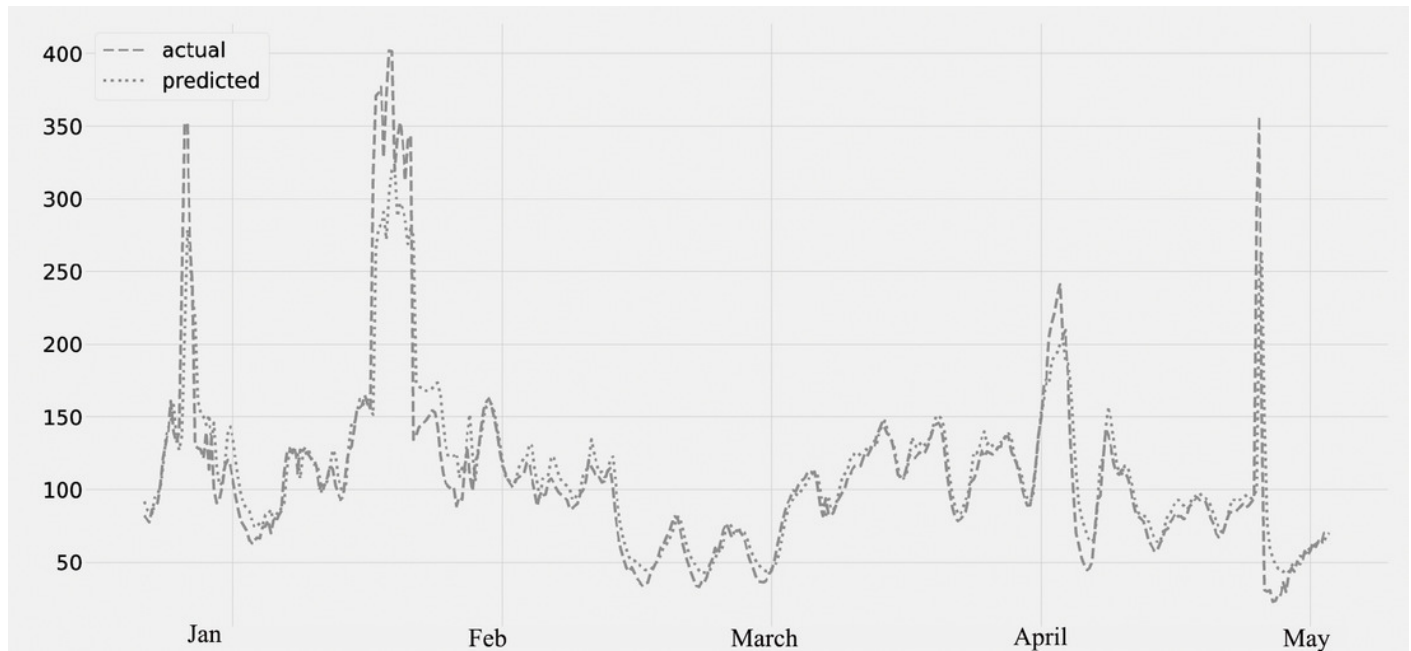



Figure 8

Time Series plot for the concentration of O3 (Jan 2020 – May 2020)

Time Series plot for the concentration of O3

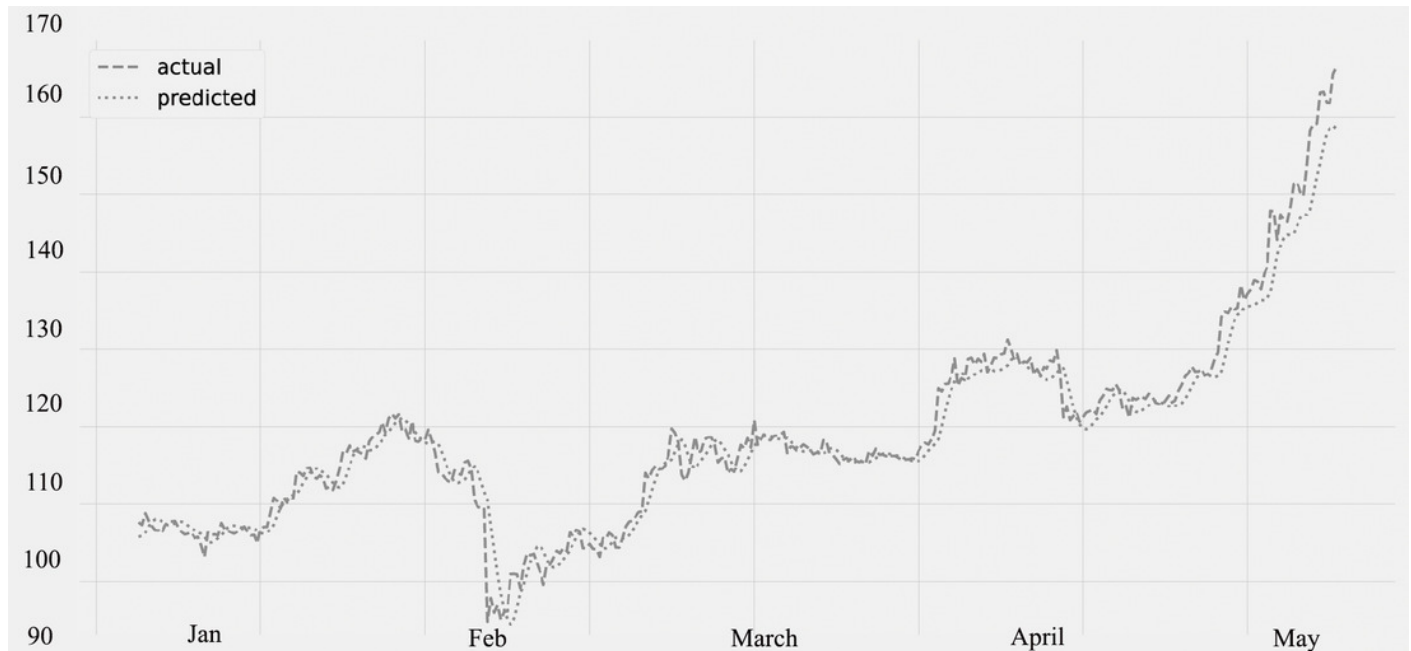



Figure 9

(a) Moving Average PM2.5, and (b) Moving average PM10

Figure 9 (a) Moving Average PM2.5 and Figure 9 (b) Moving average PM10

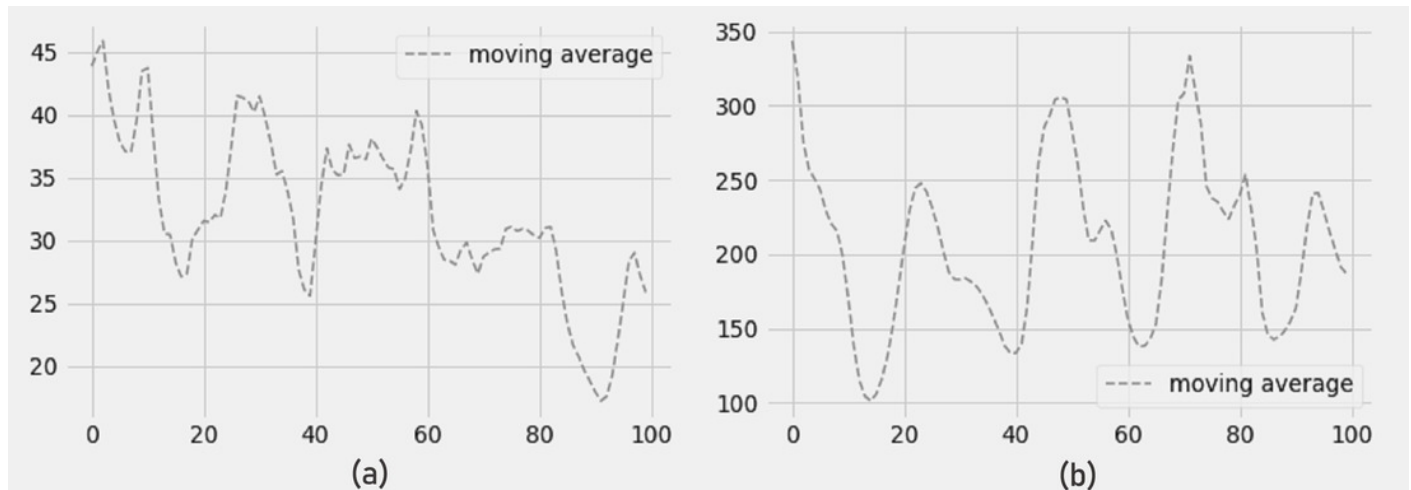



Figure 10

Moving Average of (a) NO2 concentration and (b) O3 concentration

Figure 10 (a) Moving average NO2 and Figure 10 (b) moving average O3

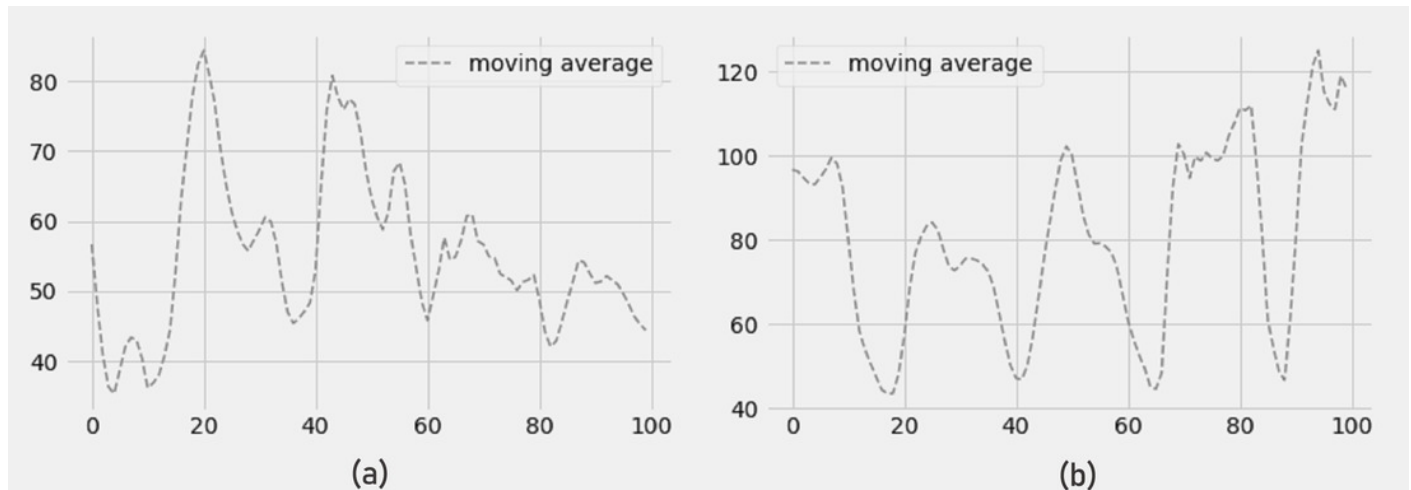
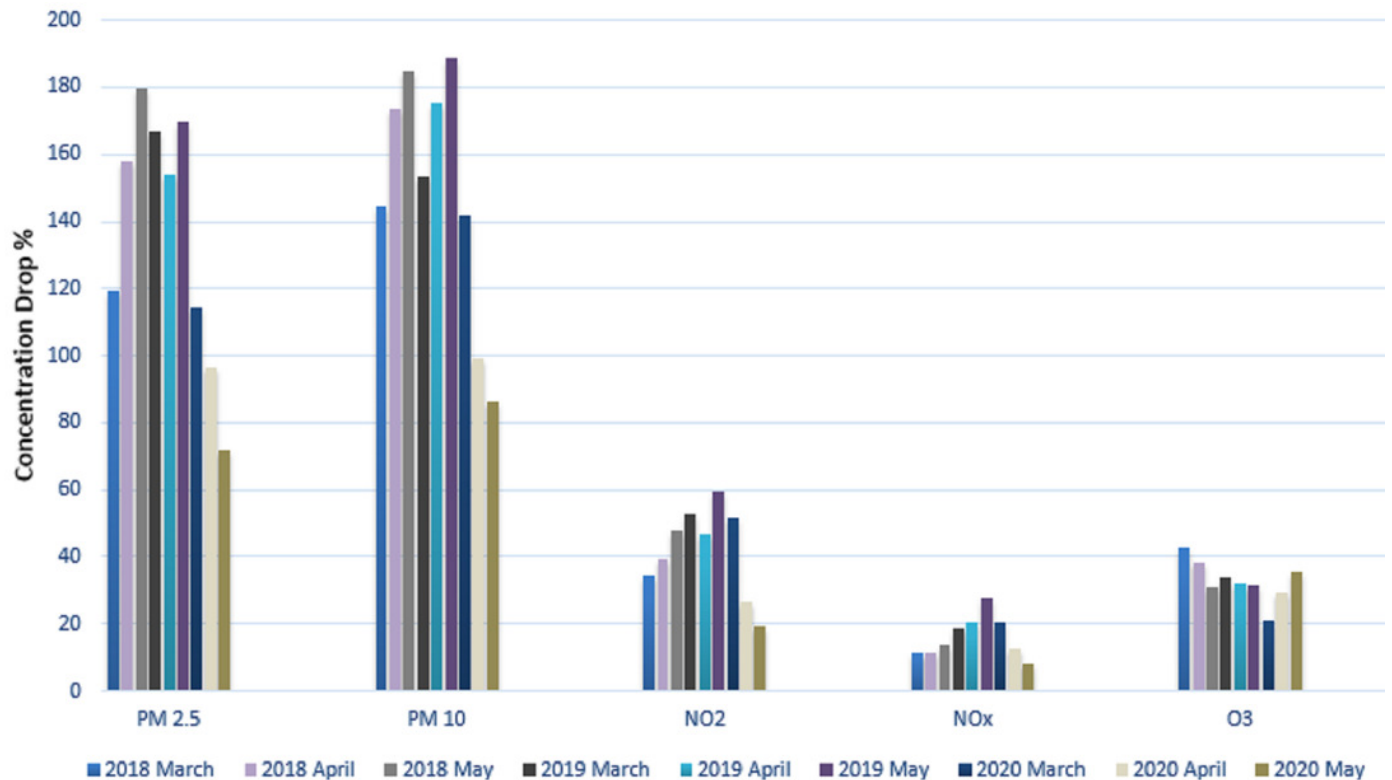



Figure 11

Average concentration of different atmospheric pollutants

Average concentration of different atmospheric pollutants

Station description for each atmospheric pollutant with the number of missing values

Weather Stations in Lahore City				
Pollutants	HAC Agri Limited	Dawn Food Head Office	Phase8-DHA	Zeenat Block
PM2.5	13	33	15	34
PM10	22	15	56	74
NO2	54	34	27	51
O3	87	98	45	18

Inferential statistical analysis for different pollutants

Pollutants	R-squared	Adj. R-squared	F-statistic	slope	intercept	P >|t|
Pm2.5	0.089	0.085	1.08	-0.0014	0.322	0.000
Pm10	0.270	0.267	3.27	-7.6301	0.269	0.000
No2	0.006	0.002	0.234	-0.008	0.2316	0.000
O3	0.052	0.048	0.00023	-0.0014	0.3056	0.000

Machine learning results for different pollutants

Machine learning results for different pollutants

Models	Pollutants	MAE	R2	RMSE
Decision Tree	PM2.5	5.327	0.283	94.607
	PM10	35.403	0.434	42.186
	NO2	6.824	0.582	154.417
	O3	17.464	0.143	716.448
SVM	PM2.5	10.752	1.375	175.17
	PM10	37.815	0.961	326.837
	NO2	5.537	0.482	102.296
	O3	26.611	1.464	118.278
Random Forest	PM2.5	8.352	1.082	153.574
	PM10	38.447	0.125	332.532
	NO2	6.076	0.105	107.886
	O3	17.714	0.194	57.125
N-BEATS	PM2.5	16.47	2.69	21.42
	PM10	18.25	4.61	19.78
	NO2	7.12	4.22	8.47
	O3	11.41	3.37	31.58
ARIMA	PM2.5	11.926	0.961	6.41
	PM10	9.241	0.76	41.809
	NO2	7.937	0.2294	32.023
	O3	4.143	0.895	36.979
CNN	PM2.5	16.54	1.25	7.57
	PM10	11.26	1.04	14.92
	NO2	8.58	0.66	11.24
	O3	9.16	0.74	39.41
LSTM	PM2.5	4.46	0.25	9.99
	PM10	8.58	0.75	12.84
	NO2	4.35	0.12	6.77
	O3	8.2	0.17	29.79

Average concentration of different pollutants during lockdown (March 2020 – May 2020) and same time frame in previous years (March – May in 2018 and March - May 2019)

Average concentration of different pollutants during lockdown (March 2020 – May 2020) and same time frame in previous years (March – May in 2018 and March - May 2019)

Average variables	2018			2019			2020		
	March	**April**	**May**	**March**	**April**	**May**	**March**	**April**	**May**
PM2.5	119.46	157.78	179.75	166.78	154.23	169.43	114.32	96.32	71.68
PM10	144.72	173.72	184.72	153.65	175.26	188.82	141.55	99.46	86.17
NO2	34.67	39.44	47.64	52.87	46.43	59.49	51.6	26.36	19.07
NOx	11.6	11.32	13.6	18.65	20.16	27.65	20.39	12.63	8.02
O3	42.72	38.23	31.09	34.03	32.4	31.72	21.11	29.1	35.5