Research Article

Air Quality Prediction Using the Fractional Gradient-Based Recurrent Neural Network

Sugandha Arora [ID],[1] **Narinderjit Singh Sawaran Singh** [ID],[2] **Divyanshu Singh** [ID],[1] **Rishi Rakesh Shrivastava** [ID],[1] **Trilok Mathur** [ID],[1] **Kamlesh Tiwari** [ID],[1] **and Shivi Agarwal** [ID][1]

[1]*Birla Institute of Technology and Science Pilani, Pilani, Rajasthan, India*
[2]*Faculty of Data Science and Information Technology, INTI International University, Persiaran Perdana BBN, Putra Nilai, 71800, Nilai, Negeri Sembilan, Malaysia*

Correspondence should be addressed to Sugandha Arora; p20180024@pilani.bits-pilani.ac.in

In this study, the air quality index (AQI) of Indian cities of different tiers is predicted by using the vanilla recurrent neural network (RNN). AQI is used to measure the air quality of any region which is calculated on the basis of the concentration of ground-level ozone, particle pollution, carbon monoxide, and sulphur dioxide in air. Thus, the present air quality of an area is dependent on current weather conditions, vehicle traffic in that area, or anything that increases air pollution. Also, the current air quality is dependent on the climate conditions and industrialization in that area. Thus, the AQI is history-dependent. To capture this dependency, the memory property of fractional derivatives is exploited in this algorithm and the fractional gradient descent algorithm involving Caputo's derivative has been used in the backpropagation algorithm for training of the RNN. Due to the availability of a large amount of data and high computation support, deep neural networks are capable of giving state-of-the-art results in the time series prediction. But, in this study, the basic vanilla RNN has been chosen to check the effectiveness of fractional derivatives. The AQI and gases affecting AQI prediction results for different cities show that the proposed algorithm leads to higher accuracy. It has been observed that the results of the vanilla RNN with fractional derivatives are comparable to long short-term memory (LSTM).

1. Introduction

With the increase in urbanization, industrialization, and traffic in the cities, the air pollutants are increasing and air quality is reducing [1]. To keep a check on the extent of air pollution, the US Environment Protection Agency has introduced a parameter called the air quality index (AQI) which tracks the daily effects of air pollutants [2]. AQI is a numerical value between 0 and 500; when the value of AQI is 0, the air quality is adequate, and if the value of AQI is 500, then the air quality is hazardous. AQI is calculated by considering major air pollutants, such as carbon monoxide (CO), nitrogen dioxide (NO_2), ozone (O_3), particulate matter (PM_{10} and $PM_{2.5}$), and sulphur dioxide

(SO_2). These pollutants are the residual gases and particles emitted from vehicles, industries, and due to climate change [3].

Biomass and coal-burning highly increase the levels of particulate matter (PM_{10} and $PM_{2.5}$) that causes haze in the air. These particles deteriorate the air composition and cause respiratory problems in living beings. Moreover, haze reduces the visibility that further affects the economic sectors such as tourism and agriculture [4]. The combustion of fossil fuels is carried out in several industries which are the main contributors of SO_2 and NO_2 in air [5, 6]. The motorized vehicles and combustion of fossil fuels also emit CO, which is another major pollutant responsible for worsening the air quality. CO is highly poisonous and can even lead to

mortality on long exposure [7]. Another major pollutant is ground-level ozone O_3, obtained from the combination of two primary pollutants, nitrogen oxides (NO_x) and volatile organic compounds (VOCs). The 95% of these primary pollutants come from oil, coal, and gasoline combustion in vehicles, industries, power plants, and households, upstream gas and oil production, combustion of residual woods, and the evaporated liquid fuels [8]. Exposure to ozone can significantly affect human health, cause asthma, and can lead to premature mortality [9]. In addition, ozone can adversely affect vegetation, damage flowers and shrubs, and reduce crop productivity [10, 11].

Air pollution is not a local phenomenon; the current quality of air is dependent on its history. The industrialization has massively impacted the environment, especially the air quality [12]. The levels of air pollutants like ground-level ozone and particulate matter are also getting influenced by modifying weather patterns that occurred due to climate change [13, 14]. The change in climate affects the temperature, humidity levels, and wind patterns, which in turn influences the air quality. In addition, the naturally occurring emissions, for example, wind-blown dust and wildfires, get provoked by climate-driven changes in meteorology that affect the air quality. The uncontrolled emission of air pollutants is gradually causing air pollution. Continuous exposure to polluted air is severely affecting human health [15] and leading to the development of lung, heart, and skin diseases [16]. Six of the world's top 10 most polluted cities are from India. Air pollution has been observed as the second biggest risk factor which is causing diseases in India and thus affecting its economy. Thus, there is a need to keep a check on air pollution in Indian cities. Each city has its unique features such as population per square km, temperature, humidity level, climate, vehicles, and industries in the region, and thus, it is better to study air pollution region-wise. Generally, the air quality in tier I and urban cities is low and it is required to give more attention in such areas.

To prevent the serious consequences of air pollution, several forecasting techniques for AQI are being developed. Based on target objectives, the techniques and approaches of forecasting are being expanded and improved. Traditional AQI forecasting techniques involve statistical techniques such as autoregressive integrated moving average (ARIMA) [17, 18], principal component regression (PCR) [19], multiple linear regression (MLR) [20], and grey models [21, 22]. These models perform well, but with the high increase in pollution, more accurate methods are required. These models are linear and thus are unable to capture the nonlinear traits [23, 24]. Even with a large amount of data, not much increase is seen in the accuracy of these models. The performances of the statistical techniques have been improved by developing hybrid techniques [25]. Artificial intelligence-based techniques are capable of analyzing the nonlinear data and thus are being recently used in the time series forecasting [26, 27]. With the availability of sufficient amount of data and computational support, AQI forecasting is being done with a deep neural network [28]. But these methods require to learn large number of parameters. Thus, a simpler and accurate method has been developed in this study using a vanilla RNN. The current level of air pollution in any area is also dependent on AQI status in the past. To capture the history dependency, fractional derivatives have been employed in the back propagation algorithm to train the vanilla RNN for the prediction of AQI in Indian cities.

In this study, five cities of different tiers are considered for AQI prediction. Bengaluru, Kolkata, and Hyderabad are tier I cities, while Patna and Talcher are tier II and tier III cities, respectively, as shown in Figure 1. The major air pollutants of Kolkata are also predicted using the proposed approach. The results show that the proposed method achieves minimum error on some fractional orders. Also, the obtained results are comparable to the LSTM. The rest of the study is structured as follows: Section 2 briefly explains the related work. The proposed approach is presented in Section 4. Section 5 discusses the experimental results obtained, followed by conclusion and future scopes in Section 6 and Section 7, respectively.

2. Preliminaries

Factional calculus is a 300-year-old branch of mathematics that deals with derivatives and integrals of noninteger order, i.e., order can be any number, be complex or real. Earlier, this domain was only theoretical involving rigorous calculations, but these derivatives are used in practical applications as well [29, 30]. Several versions of fractional derivatives and integrals have been introduced till now, where each version has unique characteristics. The most widely used versions are described.

(i) **Riemann–Liouville (RL) fractional integral operator:** This is the most frequently used version of fractional integral [31]. The α order RL fractional integral is expressed as follows:

$$_aD_t^{-\alpha} f(t) = \frac{1}{\Gamma(\alpha)} \int_a^t (t-\eta)^{\alpha-1} f(\eta)\mathrm{d}\eta. \tag{1}$$

Here, $\alpha > 0$, $t > a$, $t, a \in \mathbb{R}$, the function $\Gamma(\cdot)$ is the Gamma function, and f is a piecewise continuous function on $[0, \infty)$ and integrable on any finite subinterval of $(0, \infty)$.

(ii) **Riemann–Liouville (RL) fractional derivative:** This is the natural generalization of integer-order derivative, as this fractional derivative version and ordinary derivative are left inverse of integrals, i.e., $_a\mathbf{D}_t^{\alpha} D^{-\alpha} f(t) = D^0 f(t) = f(t)$. For $t > a$, $\alpha > 0$, and $n \in \mathbb{N}$, such that $n - 1 < \alpha < n$; the derivative of order α is evaluated by differentiating the $n - \nu$ order RL integral of function $f(t)n$ times,

$$_a\mathbf{D}_t^{\alpha} f(t) = D^n \left[_aD_t^{-(n-\alpha)} f(t) \right], \tag{2}$$

i.e.,

FIGURE 1: Map of India showing the locations of tier (I), tier II, and tier III cities represented by red, orange, and yellow color, respectively.

$$
{}_a\mathbf{D}_t^\alpha f(t) = \begin{cases} \dfrac{1}{\Gamma(n-\alpha)} \dfrac{d^n}{dt^n} \displaystyle\int_a^t \dfrac{f(\eta)}{(t-\eta)^{\gamma+1-n}} d\eta, & n-1 < \alpha < n, \\[3ex] \dfrac{d^n}{dt^n} f(t), & \alpha = n. \end{cases}
$$
(3)

It is the α order RL fractional derivative [31]. But this definition has some disadvantages as well. The most significant disadvantage is that RL derivative of order α, (< 1) of a constant is not zero.

(iii) **Caputo's fractional derivative:** For $t > a$, $\alpha > 0$, and $n \in \mathbb{N}$, such that $n-1 < \alpha < n$, the Caputo's derivative of order α is obtained by evaluating the $n-\alpha$ order RL integral of nth order derivative of function $f(t)$, i.e., ${}_a^C D_t^\alpha f(t) = {}_a D_t^{-(n-\alpha)} [D^n f(t)]$, that gives the following:

$$
{}_a^C D_t^\alpha f(t) = \begin{cases} \dfrac{1}{\Gamma(n-\alpha)} \displaystyle\int_a^t \dfrac{f^{(n)}(\eta)}{(t-\eta)^{\alpha+1-n}} d\eta, & n-1 < \alpha < n, \\[3ex] \dfrac{d^n}{dt^n} f(t), & \alpha = n. \end{cases}
$$
(4)

It is called the α order Caputo's fractional derivative. Caputo gave this definition of fractional derivative in 1967 for overcoming the limitations of RL derivative [32]. The Caputo's derivative of order

$\alpha > 0, n-1 \le \alpha < n$ for a constant c is zero. This version increased the applicability of fractional derivatives in modelling real world problems. Thus, in this study, Caputo's version of fractional derivatives has been used.

(iv) **Grünwald–Letnikov (GL) fractional derivative:** The limit definition of fractional derivative was given by Anton Karl Grünwald and Aleksey Vasilievich Letnikov in 1867 and 1868, respectively [31]. Without any assumptions on differentiability of the function for $\alpha > 0$, the α-order GL derivative of function f is expressed as follows:

$$
{}_a D_t^\alpha f(t) = \lim_{\substack{h \longrightarrow 0 \\ nh=t-a}} h^{-\alpha} \sum_{r=0}^{n} (-1)^r \binom{\alpha}{r} f(t-rh).
$$
(5)

Here, h is the step size and $\binom{\alpha}{r} = \Gamma(\alpha+1)/\Gamma(\alpha-r+1).\Gamma(r+1)$ with the Gamma function $\Gamma(\cdot)$.

Clearly, the limit definition of first-order derivative shows that evaluation of derivative involves usage of only two points. But it can be seen from limit definition of fractional derivatives and from the Caputo's definition in equation (4) and equation (5), respectively, that their evaluation involves usage of value of the function at all past points. This makes the fractional derivatives to be a nonlocal operator and incorporate memory to the systems. Due to the memory property and availability of software and other tools, fractional calculus has been used in numerous applications of science and engineering. These have been widely used in viscoelasticity [33], biology [34], signal and image processing [35], stock market [27, 36], economics [37], and in other domains with history dependency [38]. Moreover, the order of differentiation acts as a degree of freedom in the optimization process.

The nonlocality of fractional derivatives has been the major motivation for their application in different domains. Fractional calculus has been successfully used for air quality prediction [39–41]. Fractional derivative-based Kalman filter has been introduced to measure the pollutant emission and hence the air quality [39]. Several variants of fractional Kalman filters have been developed using different fractional-order derivatives version for improving the prediction accuracy [40–42]. In these air-quality models, fractional calculus is incorporated because of its long-term memory and nonlocal nature. Fractional calculus has been successfully applied in the training of artificial neural networks [43–46]. After replacing integer-order derivative by fractional-order derivative in the back propagation of the training algorithm, the update rule gets updated as follows:

$$
\Delta w = -\eta \frac{\delta^\alpha E}{\delta w^\alpha},
$$
(6)

where η and α are the learning rate and fractional order of differentiation, respectively. Chen [47] employed the fractional derivative in their backpropagation method for feedforward neural networks (FNNs) in 2013. The simulation

results showed that fractional derivative-based FNNs had a substantially better convergence speed than integer-order FNNs. The fractional derivatives have been successfully applied in the backpropagation learning algorithm of the radial basis function network [48], recurrent neural network [46], convolutional neural network [49], and even in deep neural networks and have shown significant improvement in accuracy. In our study, the effect of using fractional derivatives in the learning of the neural network has been analyzed for the prediction of air quality in few Indian cities. Earlier too, the effectiveness of fractional derivatives has been shown for nonlinear system identification, pattern classification, and Mackey–Glass chaotic time series prediction [46].

3. Fractional Gradient-Based Backward Propagation Algorithm

In this section, we introduce the fractional-order truncated backpropagation through the time algorithm on the RNN with 10 neurons in a layer. This backpropagation algorithm considers the truncated depth of the input data and the state of the network, which makes the algorithm computationally efficient. For the implementation of the backpropagation algorithm, the mean squared error at an instant is considered as follows:

$$E(s) = \frac{1}{2} \sum_{i \in \Omega} \left(\Phi(u_i(s)) - x_i(s) \right)^2 = \frac{1}{2} \sum_{i \in \Omega} \left(e_i(s) \right)^2, \quad (7)$$

where i is the output neuron, $\Phi(u_i(s))$ and $x_i(s)$ are the actual output and the expected output of the i_{th} neuron at time s, and $u_i(s) = \sum_{j \in \Omega} w_{ij} v_j(s)$ at time s, where $w_{ij}(s)$ is the weight of a signal from j_{th} neuron to i_{th} and $v_j(s)$ is the output of j_{th} neuron at time s; then, the update rule becomes as follows:

$$w_{ij}(s+1) = w_{ij}(s) - \eta \nabla^{\alpha}_{w_{ij}} E(s), \quad (8)$$

where η is the learning rate and $\nabla^{\alpha}_{w_{ij}}$ represents the factional gradient w.r.t w_{ij}. Now, $\nabla^{\alpha}_{w_{ij}} E(s)$ can be evaluated by applying the approximated chain to the error function. The actual chain rule applicable on fractional derivatives is complicated and involves special mathematical functions; thus, several approximated chain rules have been developed for fractional derivatives. [50–53] The chain rule given by expression (14) has been obtained by using fractional Taylor's series expansion for differentiable function. Consider a differentiable function, say f then for a small h,

$$f(x+h) = f(x) + \frac{h^{\alpha}}{\Gamma(1+\alpha)} f^{(\alpha)}(x) + \frac{h^{2\alpha}}{\Gamma(2+\alpha)} f^{(2\alpha)}(x)$$

$$+ \frac{h^{3\alpha}}{\Gamma(3+\alpha)} f^{(3\alpha)}(x) + \cdots. \quad (9)$$

Then,

$$\frac{\Delta f(x)}{h^{\alpha}} = \frac{1}{\Gamma(1+\alpha)} f^{(\alpha)}(x) + \frac{h^{\alpha}}{\Gamma(2+\alpha)} f^{(2\alpha)}(x)$$

$$+ \frac{h^{2\alpha}}{\Gamma(3+\alpha)} f^{(3\alpha)}(x) + \cdots. \quad (10)$$

Taking limit $h \longrightarrow 0$, we get

$$\lim_{h \longrightarrow 0} \frac{\Delta f(x)}{h^{\alpha}} = \frac{1}{\Gamma(1+\alpha)} f^{(\alpha)}(x). \quad (11)$$

Also, $\lim_{h \longrightarrow 0} \Delta^{\alpha} f(x)/h^{\alpha} = f^{(\alpha)}(x)$,

$$\Rightarrow \Delta^{\alpha} f(x) \approx \Gamma(1+\alpha) \Delta f(x) 0 < \alpha < 1. \quad (12)$$

From above equation, we can also say $d^{\alpha} f(x) \approx \Gamma(1+\alpha) df(x) 0 < \alpha < 1$. Using this result,

$$\frac{d^{\alpha} f(v(x))}{dx^{\alpha}} = \frac{d^{\alpha} f(u(x))}{d^{\alpha} v} \frac{d^{\alpha} v}{dx^{\alpha}} = \frac{\Gamma(\alpha+1) df(v(x))}{\Gamma(\alpha+1) dv} v_x^{(\alpha)}(x)$$

$$= g_u'(v) . v_x^{(\alpha)}. \quad (13)$$

Hence,

$$D_x^{\alpha} f(v(x)) = D_u' f(v) . D_x^{\alpha} v(x). \quad (14)$$

After using the abovementioned fractional chain rule, we get

$$\nabla^{\alpha}_{w_{ji}} E(s) = \frac{\partial E(s)}{\partial e_j(s)} \cdot \frac{\partial e_j(s)}{\partial v_j(s)} \cdot \frac{\partial v_j(s)}{\partial u_i(s)} \cdot \frac{\partial^{\alpha} u_i(s)}{\partial w_{ji}^{\alpha}(s)}. \quad (15)$$

Suppose $\partial E(s)/\partial u_j(s) = \delta_j(s)$, then

$$\frac{\partial E(s)}{\partial e_j(s)} \cdot \frac{\partial e_j(s)}{\partial v_j(s)} \cdot \frac{\partial v_j(s)}{\partial u_i(s)} = \frac{\partial E(s)}{\partial u_j(s)} \cdot \frac{\partial^{\alpha} u_i(s)}{\partial w_{ji}^{\alpha}(s)} = \delta_j(s) \cdot \frac{\partial^{\alpha} u_i(s)}{\partial w_{ji}^{\alpha}(s)}. \quad (16)$$

Now, as in this study Caputo's version of fractional derivative is being used and $^{C}_{0}D_x^{\alpha} x^p = \Gamma(p+1) x^{p-\alpha} / \Gamma(p-\alpha+1)$, for $p > -1$, then the following holds:

$$\frac{\partial^{\alpha} u_i(s)}{\partial w_{ji}^{\alpha}(s)} = \frac{w_{ji}^{1-\alpha}(s) v_i(s)}{\Gamma(2-\alpha)}. \quad (17)$$

Thus, from equations (4), (16), and (17), the following final update rule is obtained:

$$w_{kj}(s+1) = w_{kj}(s) - \eta \sum_{s=n-h+1}^{n} \delta_k(s) \frac{w_{kj}^{1-\alpha}(s) v_j(s)}{\Gamma(2-\alpha)}. \quad (18)$$

4. Proposed Approach

In this study, the vanilla RNN has been employed to predict the AQI value of a day based on the previous sequential AQI data of five different cities. RNNs are capable of learning the sequential pattern of historical data. Furthermore, the accuracy of the system has been improved by incorporating memory into the system using the fractional gradient descent algorithm.

4.1. Data Exploration and Processing. In this study, the continuous AQI data have been used to predict future unseen AQI values. For each city, a continuous-time patch of around 1000 data points has been used from the AQI dataset. The sample data can be seen in Table 1. For constructing the training data, a min-max scaler has been used to scale the data values between 0 and 1. The predicted values are also obtained between 0 and 1, which are inverse-transformed later to evaluate the final predicted AQI value.

4.2. Neural Network Architecture and Training. The vanilla RNN has been used in the proposed model which has a single-layered architecture with 10 nodes in it. The fractional gradient-based RNN model is built from scratch using the NumPy library, and the Pandas library is used for data preprocessing. Forward propagation and fractional gradient-based back propagation as given by (18) have been used for training the vanilla RNN. On the other hand, for the LSTM, the TensorFlow library is used to produce all the results, and the integer-order gradient descent algorithm is used to train the model. Backpropagation has been used in both the models through 10 days (timestamps) which would predict the AQI data for the next (11th) day. Training and testing sets have 600–800 and 100 data values, respectively. For the initializing of weights for the models, Xavier's initialization has been used with 0.1 learning rate; 80 epochs were used to train the model for each city and each fractional order. Figure 2 shows the architecture of the RNN with fractional gradient-based backpropagation.

4.3. Evaluation Parameter. The standard evaluation metrics for forecasting models *viz* root mean squared error (RMSE) and mean absolute percentage error (MAPE) have been employed to assess the performance of the proposed model in the prediction of AQI of different Indian cities and the major pollutants in one of those cities. The lesser the value of RMSE and MAPE, the better the performance of the predictor. These errors measure the performance of forecasting, climatology, and regression analysis for verifying the experimental results. The detailed information related to these parameters is provided.

The root mean square error (RMSE) is the square root of the average of the squared difference between the actual and predicted value. RMSE can be expressed by the following expression:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N} (\widehat{o}_t - o_t)^2}{N}}, \tag{19}$$

where $\widehat{o}_t$ is the predicted value and o_t is the expected output for iteration t, which are observed for N times.

The mean absolute percentage error (MAPE) is the average percentage of the absolute difference between the actual and predicted values divided by the actual value for each time period [54]. MAPE can be expressed by the following expression:

$$MAPE = \frac{100}{N} \sum_{t=1}^{N} \frac{|\widehat{o}_t - o_t|}{|o_t|}, \tag{20}$$

where $\widehat{o}_t$ is the predicted value and o_t is the expected output for iteration t, which are observed for N times. This is a form of percentage error, which has helped in the analysis of the proposed model in different situations.

5. Results and Analysis

This section describes the AQI dataset of five cities chosen, the results obtained by using the proposed approach on the AQI data, and the discussion of comparison between predictions of LSTM and the proposed approach on different fractional orders. The performance of networks is measured using RMSE and MAPE.

5.1. Dataset. The AQI dataset of five cities for 2015–2020 is considered which is publicly available at the official portal of the Central Pollution Control Board, Government of India (http://cpcb.nic.in/). The dataset consists of daily air quality levels at various stations across multiple cities in India which are obtained by averaging out the hourly value of AQI. Indian cities chosen for the analysis includes Kolkata, Hyderabad, Bengaluru, Patna, and Talcher. Basic information related to these cities is as follows:

(i) **Kolkata** ($22°34'03''$ N $88°43'57''$ E), located in West Bengal, is a tier I and the seventh most populous city of India with third-most populous metropolitan area. The concentration of pollutants such as sulphur dioxide and nitrogen dioxide remains within the limit, but the presence of particulate matter in air is high and is increasing over the years. Due to this, air pollution is severe and is causing respiratory ailments such as lung cancer.

(ii) **Bengaluru** ($12°58'44''$ N $77°35'30''$ E), located in Karnataka, is also a tier I and the third most populous city of India with fifth most populous metropolitan area. Bengaluru is also considered as "Silicon Valley of India" because it is the nation's top IT exporter. This IT hub region is the most polluted and is causing several environmental issues. Due to the large population, Bengaluru generates tonnes of solid waste which is polluting the environment. Thus, the large population and IT hub of Bengaluru is the major reason for air pollution.

TABLE 1: Data samples of all cities.

|			Bengaluru			
Date	2019-09-01	2019-09-02	2019-09-03	2019-09-04	2019-09-05	2019-19-06	2019-19-07
AQI	61	48	51	51	56	64	63
			Kolkata			
Date	2018-06-16	2018-06-17	2018-06-18	2018-06-19	2018-06-20	2018-06-21	2018-06-22
AQI	119	113	107	148	94	100	60
			Hyderabad			
Date	2017-11-01	2017-11-02	2017-11-03	2017-11-04	2017-11-05	2017-11-06	2017-11-07
AQI	150	156	158	114	91	80	94
			Patna			
Date	2017-11-05	2017-11-06	2017-11-07	2017-11-08	2017-11-09	2017-11-10	2017-11-11
AQI	276	289	286	354	430	439	429
			Talcher			
Date	2018-02-08	2018-02-09	2018-02-10	2018-02-13	2018-02-14	2018-02-15	2018-02-16
AQI	311	321	343	343	269	243	350
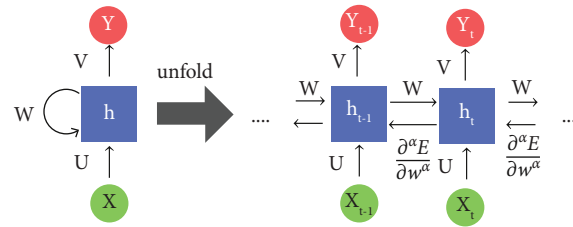
FIGURE 2: RNN architecture with fractional backpropagation.

(iii) **Hyderabad** ($17°21'42''$ N $78°28'29''$ E), located in Telangana, is also a tier I and the fourth most populous city of India with sixth most populous metropolitan area. Again, due to the large population, increased economic activity, and rapid urbanization, tonnes of solid waste are generated, and disposal of such waste becomes hazardous and pollutes the environment. The particulate matter (PM_{10}) dispersed in the atmosphere causes around 2500 deaths each year.

(iv) **Patna** ($25°36'0''$ N $85°6'0''$ E), located in Bihar, is a tier II city with a high population. Air pollution is a major issue in this city. The situation in winter becomes even worse due to dense smog, leading to an increase in mortality. Patna was declared as the second most air polluted city in India, in the WHO survey of 2014.

(v) **Talcher** ($20°57'0''$ N $85°13'48''$ E), located in Angul district of Orissa, is a tier III city. This is a small city with less population, but Talcher has the country's biggest coalfield with the highest coal reserve of around 52 billion tonnes. The presence of these coal mines leads to air pollution.

The cities of different tiers have been chosen where air pollution is a major issue. To summarise, the cities with a large population or with a high emission rate of air pollutants affecting human health are considered. Around 600 normalized data points for each city have been used for the analysis, which is divided into train and test data in the ratio of 4 : 1. The model has been tested on the data for 100 days for each city.

5.2. AQI Prediction Results Using Fractional-Order Gradient Learning. The performance of the vanilla RNN in predicting AQI values of each city using the fractional backpropagation algorithm has been analyzed. To assess the performance of the model, RMSE and MAPE are computed. The prediction performance of the RNN using the fractional gradient descent algorithm with values of fractional orders in the neighborhood of 1 is compared with the performance of the RNN with the traditional integer-order gradient descent algorithm where the order remains 1. The values of fractional order α which are considered are 6/9, 7/9, 8/9, 1, 10/9, and 11/9 [49]. The results obtained at different orders using the proposed approach can be seen in Table 2. The graphs in Figures 3–7 show the comparison between actual and predicted output for Bengaluru, Kolkata, Hyderabad, Patna, and Talcher, respectively. In all the graphs, the expected output and the actual output are represented by the yellow lines and blue lines, respectively. It can be observed that the least RMSE and MAPE are acquired by the vanilla RNN at some fractional orders, either on 7/9 or 8/9 for all the cities. The model achieved minimum RMSE and MAPE of 13.22 and 06.11%, respectively, at $\alpha = 8/9$ for Bengaluru. Also, the minimum RMSE and MAPE are found to be 19 and 7.02% for Kolkata and 23.85 and 7.43% for Patna at $\alpha = 8/9$.

TABLE 2: Performance comparison of the fractional gradient-based RNN with the integer gradient-based RNN and LSTM.

Model	Bengaluru	Kolkata	Hyderabad RMSE	Patna	Talcher	Bengaluru	Kolkata	Hyderabad MAPE (%)	Patna	Talcher
$\alpha = 6/9$	21.47	31.35	25.35	44.30	14.25	14.41	18.21	16.67	14.78	06.62
$\alpha = 7/9$	19.36	26.12	**07.41**	34.08	**11.40**	11.32	11.58	**03.22**	10.64	**04.15**
$\alpha = 8/9$	**13.22**	**19.00**	09.75	**23.85**	29.64	**06.11**	07.02	05.92	**07.43**	16.30
$\alpha = 1$	14.78	21.37	08.58	25.44	15.96	07.15	09.31	05.23	08.07	06.11
$\alpha = 10/9$	24.99	28.50	11.31	38.56	21.09	16.60	12.90	05.49	11.51	06.66
$\alpha = 11/9$	35.20	21.85	09.95	29.53	24.51	22.54	10.50	04.99	12.90	07.90
LSTM	**10.78**	**19.00**	**07.02**	24.12	13.11	**05.78**	**06.61**	03.45	07.83	**04.11**
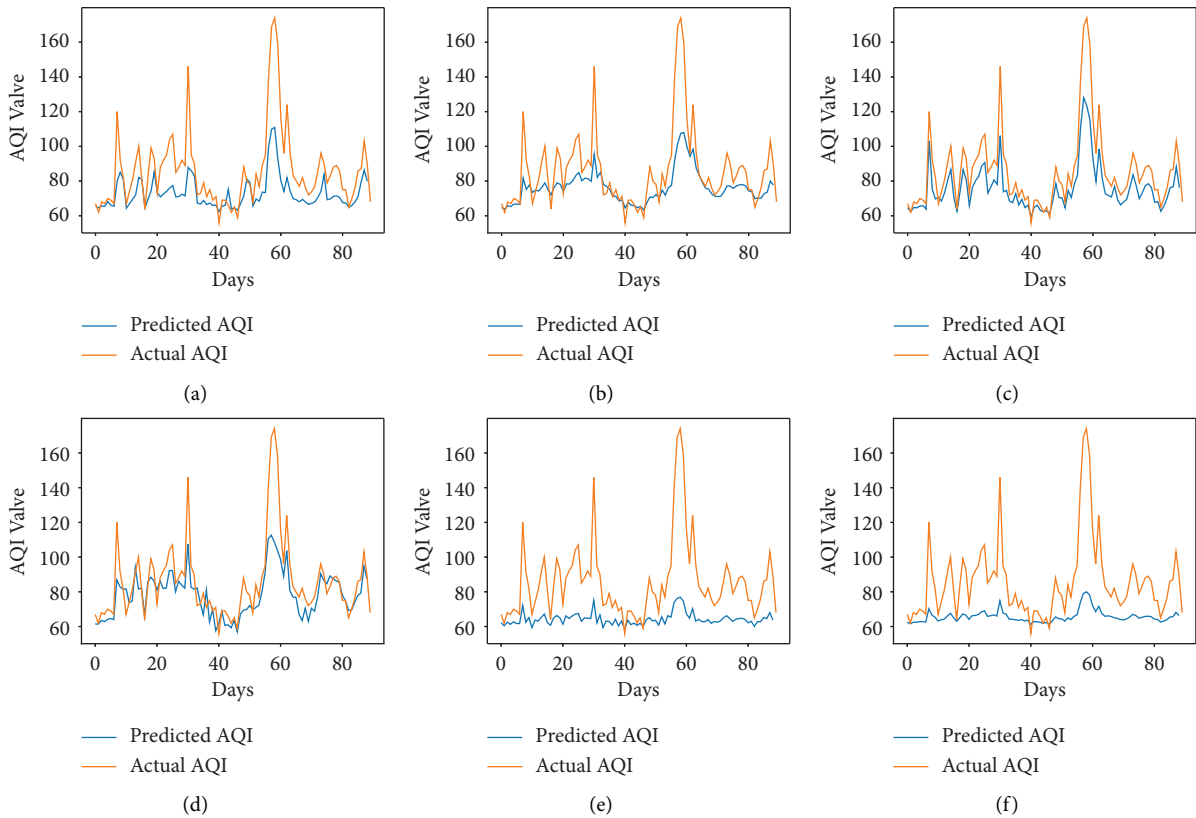
FIGURE 3: Predicted vs. actual AQI of Bengaluru city for the period of 100 days from 29-08-2019 to 07-12-2019. (a) $\alpha = 6/9$. (b) $\alpha = 7/9$. (c) $\alpha = 8/9$. (d) $\alpha = 1$. (e) $\alpha = 10/9$. (f) $\alpha = 11/9$.

Moreover, the minimum RMSE and MAPE are found to be 7.41 and 3.22% for Hyderabad and 11.40 and 4.15% for Talcher at $\alpha = 7/9$. Thus, it can be concluded from the results that the fractional-order gradient is more accurate than the integer-order gradient algorithm. Moreover, the proposed model performed best for $\alpha = 0.7$ by achieving the least MAPE of 3.22% for Hyderabad among all the cities.

5.3. Comparison of Results Obtained by the Proposed Method and LSTM.
The prediction of AQI for the same set of datasets of all cities has been done using LSTM as well with the same number of timestamps, nodes, and the same procedure for inputs as done for the fractional RNN. The obtained results are also shown in Table 2. It can be concluded from the table that the performance of the LSTM is comparable to the vanilla RNN with fractional gradient learning. The RMSE values obtained by fractional gradient learning on Vanilla RNN are 23.85 and 11.40 which are lesser than 24.12 and 13.11 RMSE achieved by LSTM, corresponding to Talcher and Patna, respectively. In addition, the RMSE of the fractional-based RNN and LSTM is found to be 19 for Kolkata which is equal. Similarly, the MAPE for Hyderabad and Patna is found to be 03.22% and 7.43%, respectively, by the vanilla RNN with fractional gradient learning, whereas higher MAPE of 3.45% and 7.83% for Hyderabad and Patna, respectively, is achieved by LSTM. For Talcher, MAPE is found to be 4.15% corresponding to the fractional gradient-based RNN and is found to be 4.11% for LSTM which are almost equivalent. It can be
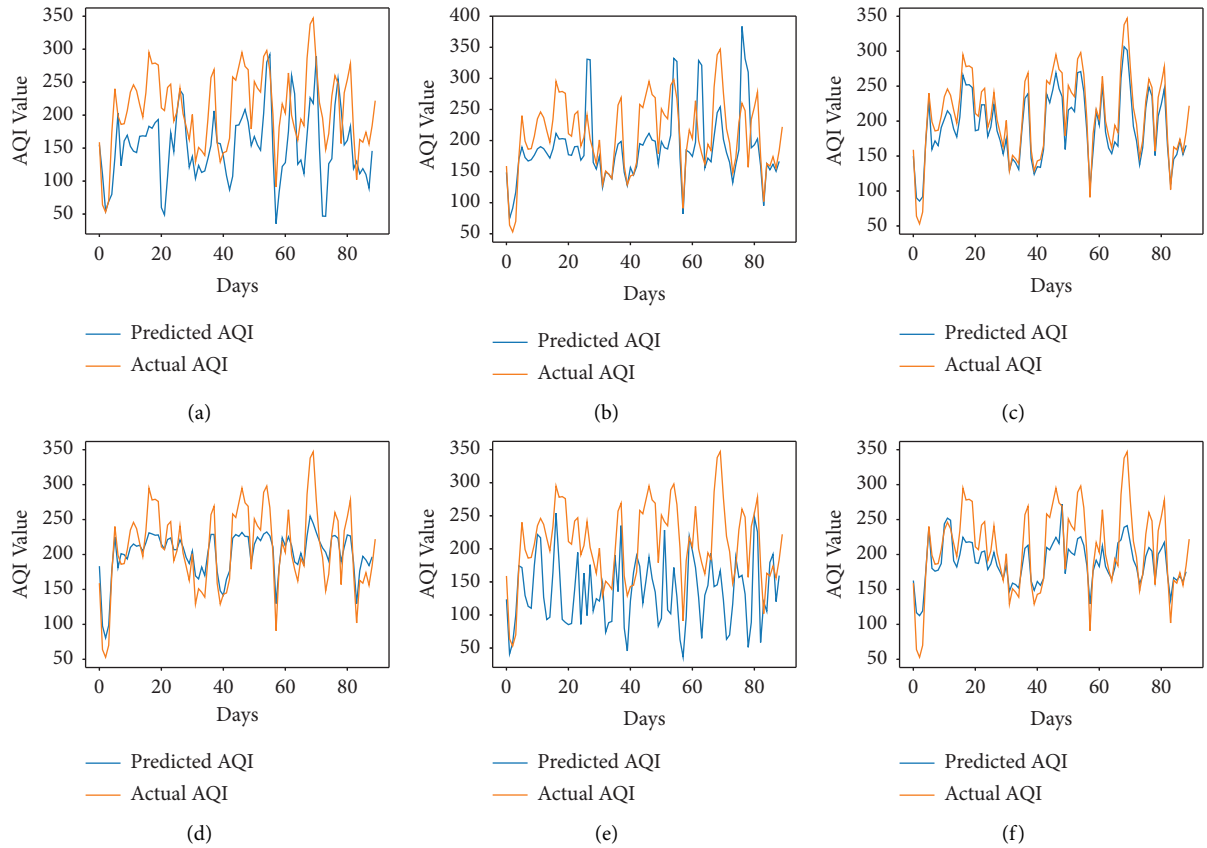
FIGURE 4: Predicted vs. actual AQI of Kolkata city for the period of 100 days from 28-10-2019 to 04-02-2020. (a) $\alpha = 6/9$. (b) $\alpha = 7/9$. (c) $\alpha = 8/9$. (d) $\alpha = 1$. (e) $\alpha = 10/9$. (f) $\alpha = 11/9$.
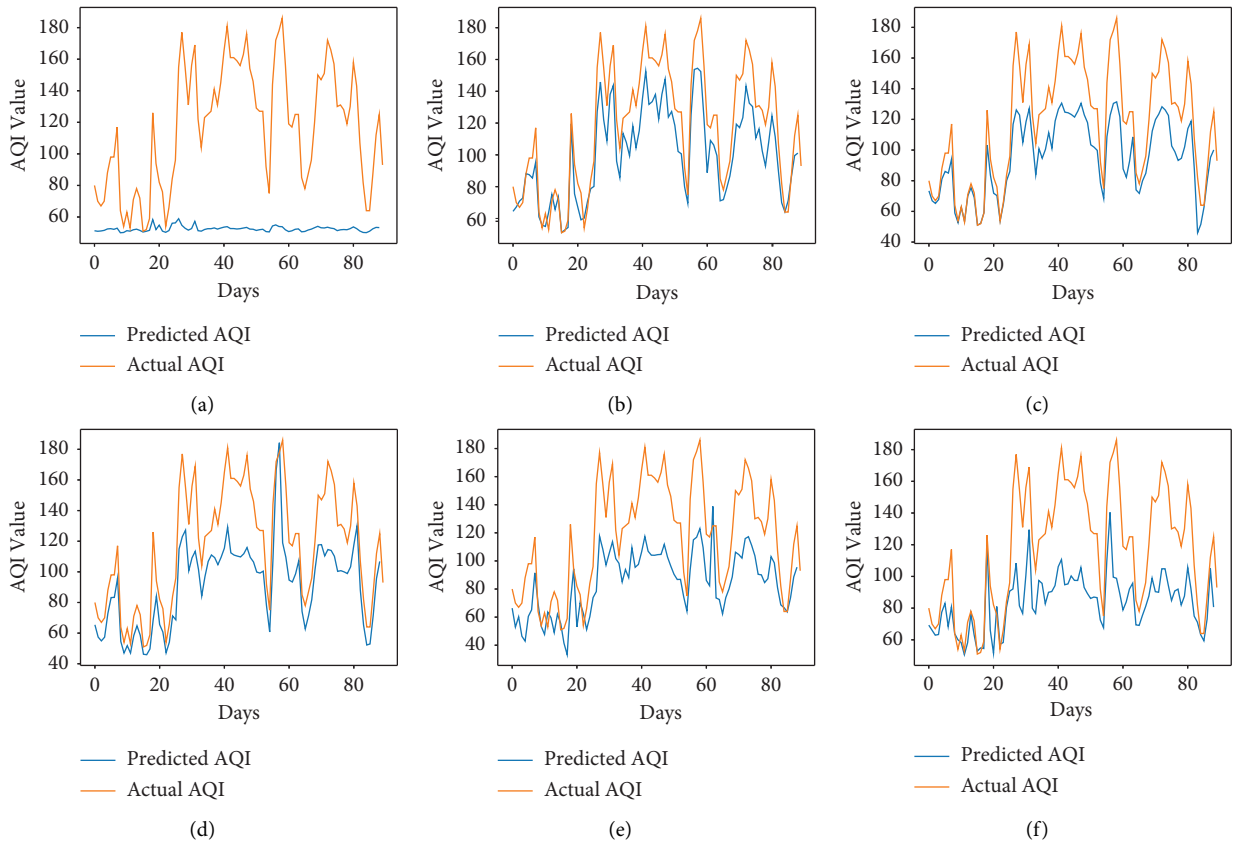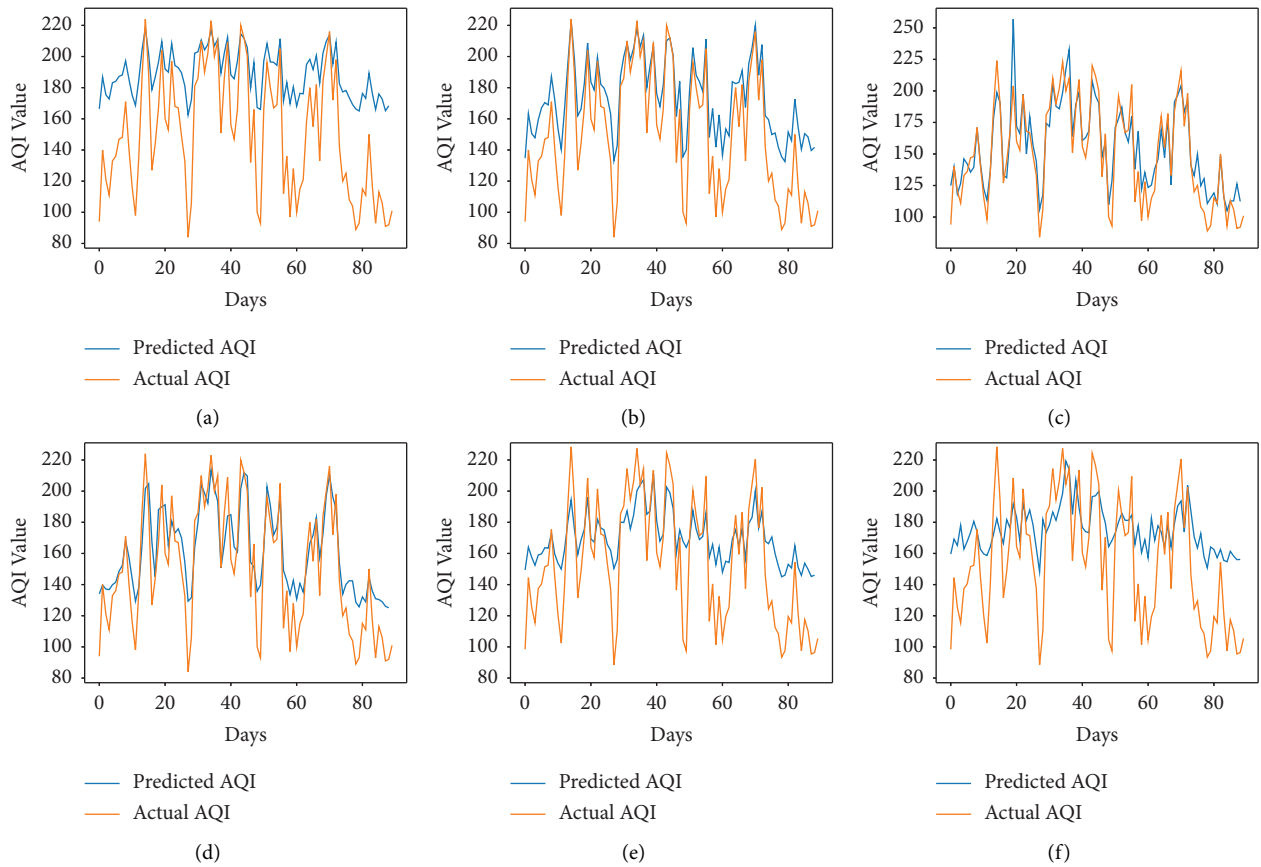

FIGURE 5: Predicted vs. actual AQI of Hyderabad city for the period of 100 days from 29-09-2019 to 06-01-2020. (a) $\alpha = 6/9$. (b) $\alpha = 7/9$. (c) $\alpha = 8/9$. (d) $\alpha = 1$. (e) $\alpha = 10/9$. (f) $\alpha = 11/9$.

Figure 6: Predicted vs. actual AQI of Patna city for the period of 100 days from 26-03-2019 to 04-07-2019, (a) $\alpha = 6/9$. (b) $\alpha = 7/9$. (c) $\alpha = 8/9$. (d) $\alpha = 1$. (e) $\alpha = 10/9$. (f) $\alpha = 11/9$.
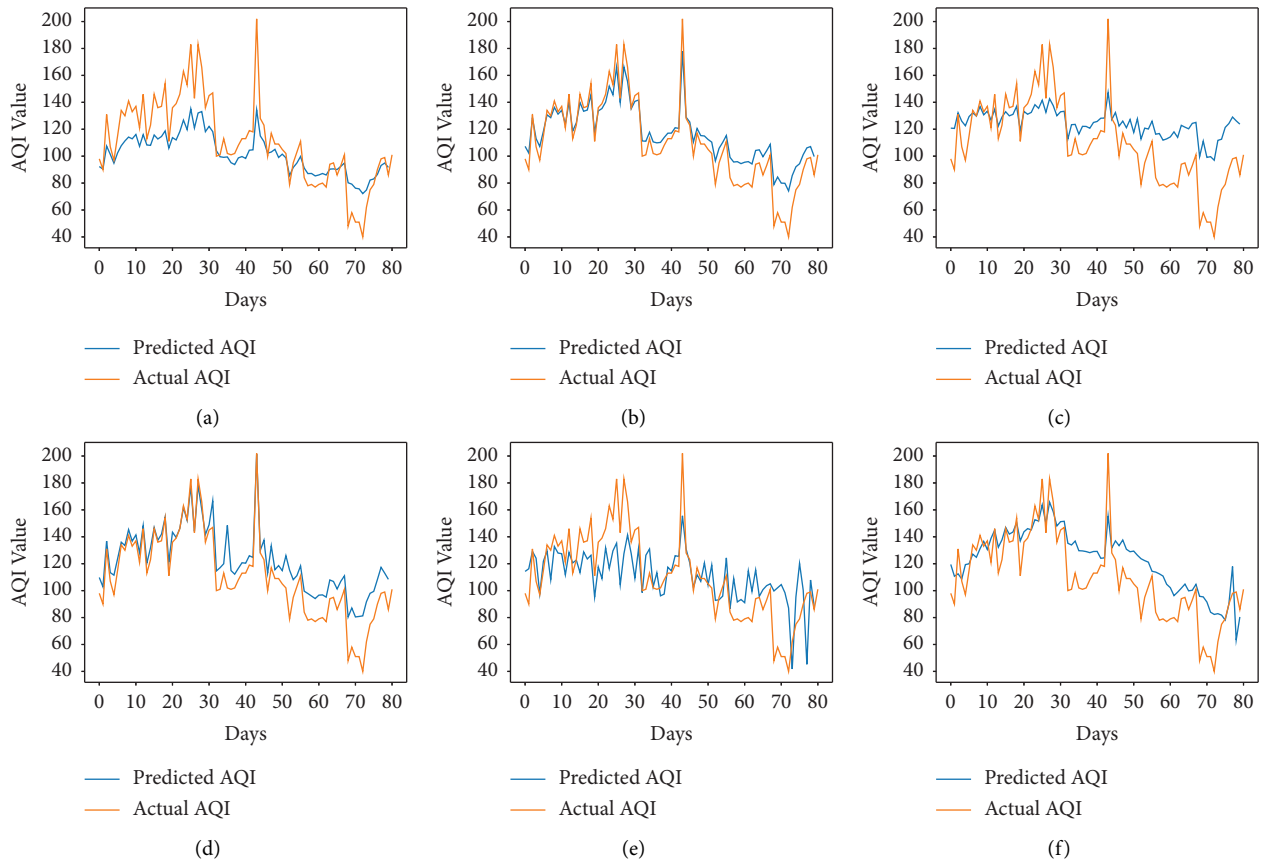

Figure 7: Predicted vs. actual AQI of Telchar city for the period of 100 days from 01-03-2020 to 08-6-2020. (a) $\alpha = 6/9$. (b) $\alpha = 7/9$. (c) $\alpha = 8/9$. (d) $\alpha = 1$. (e) $\alpha = 10/9$. (f) $\alpha = 11/9$.
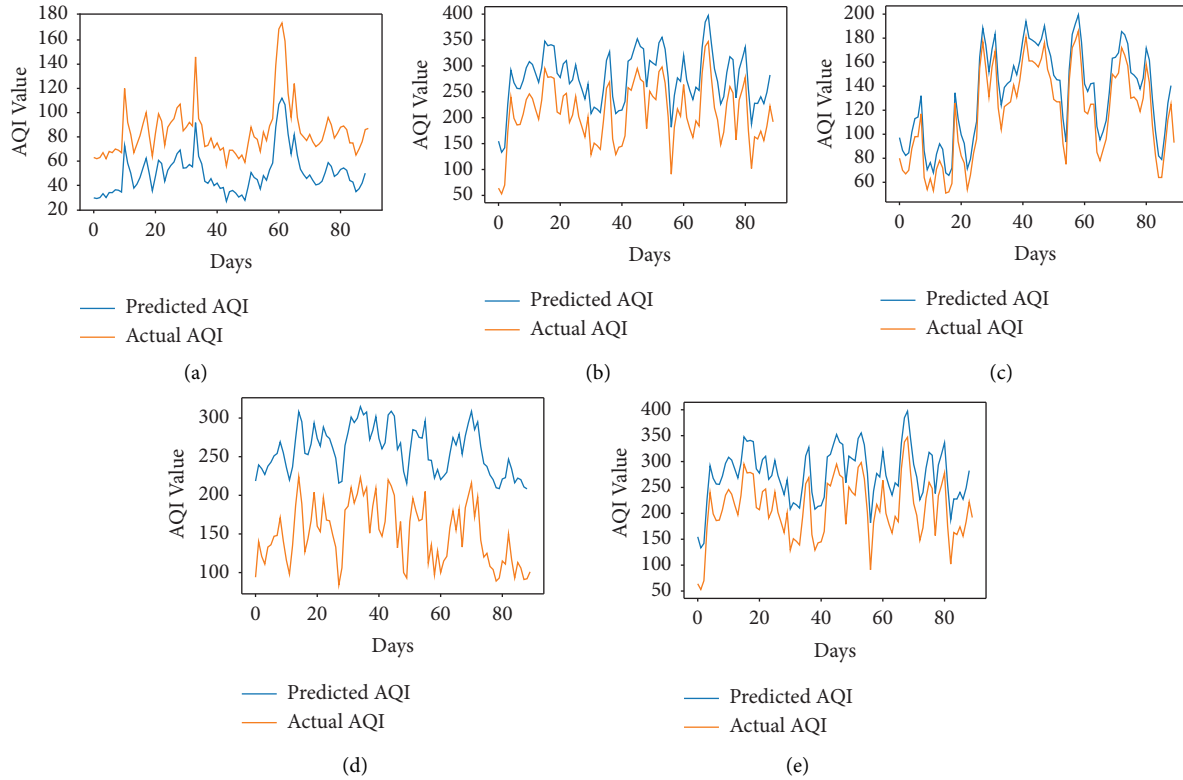
(a)

(b)

(c)

(d)

(e)

FIGURE 8: Predicted vs. actual AQI of all cities for 100 days using LSTM, (a) Bengaluru. (b) Kolkata. (c) Hyderabad. (d) Patna. (e) Telchar.
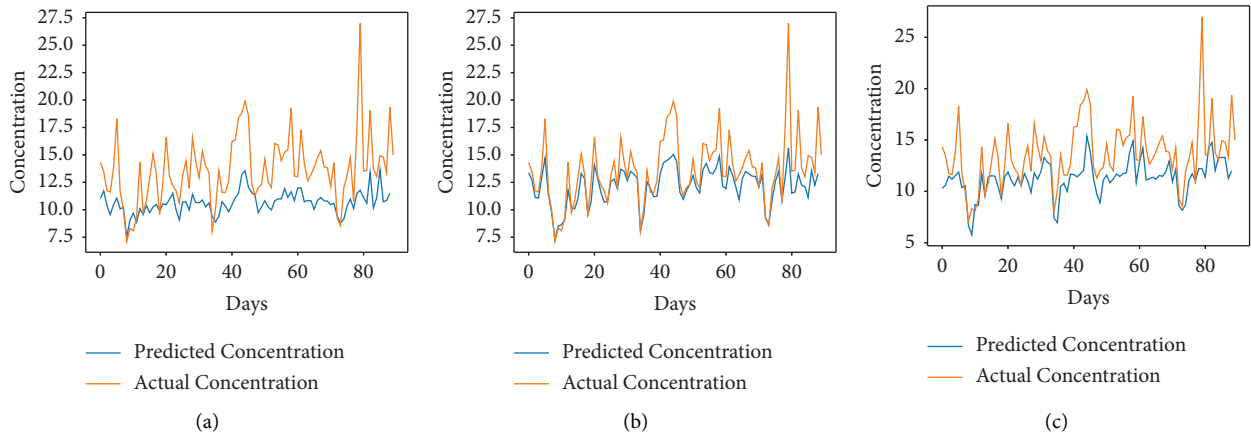

(a)

(b)

(c)

FIGURE 9: Predicted vs. actual sulphur dioxide (SO_2) in air in Kolkata for the period of 100 days from 28-10-2019 to 04-02-2020. (a) $\alpha = 7/9$. (b) $\alpha = 8/9$. (c) $\alpha = 1$.

seen that the least MAPE of 3.22% is obtained by the fractional gradient-based RNN for Hyderabad as compared to other cities and models. Figure 8 shows the comparison between the expected AQI value and the AQI value predicted for all the cities by LSTM.

5.4. Prediction Results of Major Pollutants in Kolkata Using Different Fractional-Order Gradient Learning.
The proposed approach has been implemented for the prediction of the

concentration of major pollutants such as SO_2, CO, and PM_{10}. Here, the considered time is also the same as used in the prediction of the AQI of Kolkata. As we have seen in the above section, the performance of the algorithm is found to be better either for $\alpha = 7/9$ or 8/9. So, we have considered these two fractional values to compare the results with the integer-order-based learning of the vanilla RNN. Figures 9–11 show the comparison between expected air pollutant concentrations and actual concentrations in Kolkata. It can be observed from Table 3 that minimum
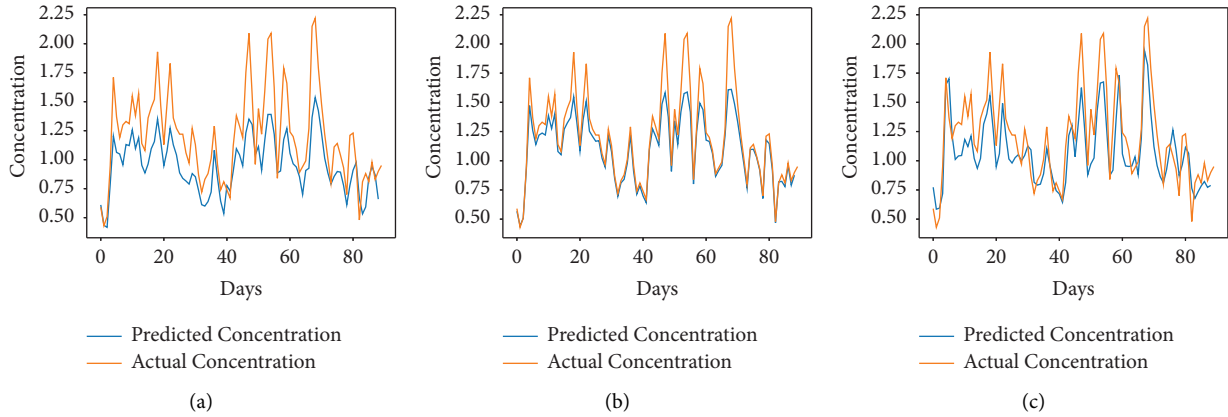
FIGURE 10: Predicted vs. actual carbon monoxide (CO) in air in Kolkata for the period of 100 days from 28-10-2019 to 04-02-2020. (a) $\alpha = 7/9$. (b) $\alpha = 8/9$. (c) $\alpha = 1$.
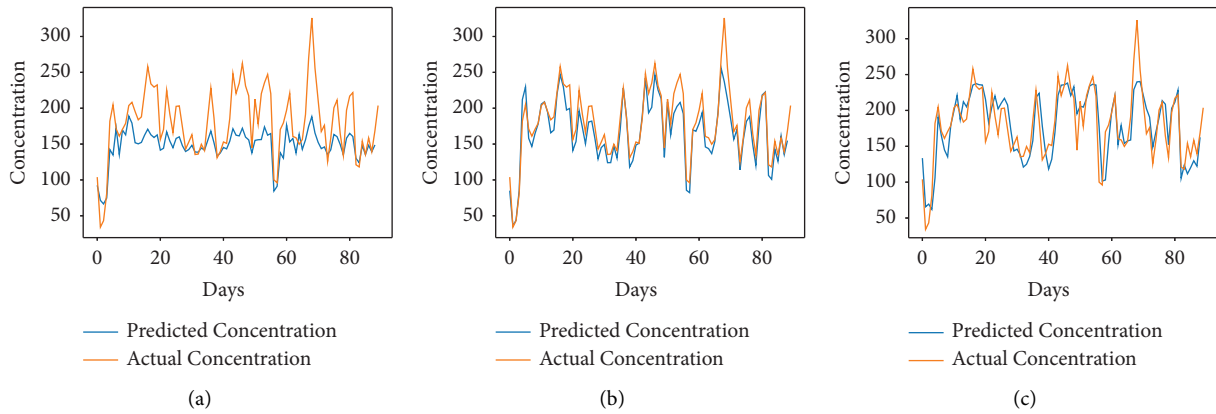

FIGURE 11: Predicted vs. actual particulate matter (PM_{10}) in air in Kolkata for the period of 100 days from 28-10-2019 to 04-02-2020. (a) $\alpha = 7/9$. (b) $\alpha = 8/9$. (c) $\alpha = 1$.

TABLE 3: RMSE corresponding to the fractional gradient-based RNN for different orders in predicting concentration of different pollutants responsible for air pollution in Kolkata city.

Pollutant	RMSE			MAPE (%)		
	$\alpha = 7/9$	$\alpha = 8/9$	$\alpha = 1$	$\alpha = 7/9$	$\alpha = 8/9$	$\alpha = 1$
SO_2	1.9	**1.62**	1.77	7.00	**5.55**	6.83
CO	0.20	**0.12**	0.15	5.21	**4.20**	4.89
PM_{10}	36.45	**19.25**	33.75	10.95	**6.77**	11.21

RMSE and MAPE for each city are attained at order $\alpha = 8/9$, thus outperforming the traditional integer-order learning for the vanilla RNN. Moreover, the least MAPE of 4.70% is achieved in the prediction of CO, and thus, the proposed model is better for predicting the concentrations of CO as compared to other pollutants.

6. Conclusion

In this study, the fractional-order gradient has been used in the backpropagation of the vanilla RNN for the AQI prediction of five Indian cities of all tiers. The proposed approach has been used for the prediction of major air pollutants in tier I Kolkata city. Through the results of

prediction of AQI of multiple cities and prediction of air pollutants, it has been observed that the minimum error on predictions is achieved at a fractional order. Most cities achieve better results when the order is equal to 8/9. The architecture of the vanilla RNN is much simpler than the structure/functioning of an LSTM, but the predictions made by RNNs with fractional gradient-based backpropagation are comparable and sometimes even better than LSTM with the integer-order gradient descent algorithm. Achieving lesser RMSE and MAPE with simpler architecture shows the effectiveness of fractional gradient over integer-order gradient descent. The least MAPE value is found for Hyderabad by using the fractional gradient-based RNN as compared to other cities and models. In addition, the least MAPE is

achieved during predictions of CO concentrations in Kolkata. Therefore, the proposed model is better in the prediction of AQI values of Hyderabad as compared to other cities and CO concentrations in the air of Kolkata as compared to other air pollutants. From the results, it can be seen that RMSE is more for Kolkata and Patna. Patna is even amongst the world's top 10 most polluted cities, and particulate matter is increasing in Kolkata each year. Hence, the memory property of fractional derivatives can be well exploited with deep neural networks for dealing with more complex and dynamic data.

7. Future Scope

This study can be extended by predicting other air pollutants in the cities, and from there, AQI values can also be predicted. Using this strategy, major air pollutants in a city can be detected and stringent actions can be taken accordingly to prevent further damage. A portfolio of economic activities can be created considering the air quality of the particular city and also detecting the most affecting gases among them in the future. The order of the derivative is chosen manually in this study, due to which results are evaluated only on a few values of order. Hence, there is a need to develop an adaptive method that automatically evaluates the optimal order for a particular city or a set of data. Methods like particle swarm optimization (PSO) and genetic algorithms can be employed for optimizing the order of differentiation. Fractional gradient descent can be used with suitable architectures for different cities. Through the results of predictions of various gases of the city, we can find a better way to develop in a sustainable way.

Data Availability

The AQI dataset of five cities for 2015–2020 is considered, which is publicly available at https://cpcb.nic.in, the official portal of the Central Pollution Control Board, Government of India.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. Ma, Y. Ding, J. C. Cheng et al., "Identification of high impact factors of air quality on a national scale using big data and machine learning techniques," *Journal of Cleaner Production*, vol. 244, Article ID 118955, 2020.

[2] G. K. Kang, J. Z. Gao, S. Chiao, S. Lu, and G. Xie, "Air quality prediction: big data and machine learning approaches," *International Journal of Environment and Sustainable Development*, vol. 9, no. 1, pp. 8–16, 2018.

[3] L. Bai, J. Wang, X. Ma, and H. Lu, "Air pollution forecasts: an overview," *International Journal of Environmental Research and Public Health*, vol. 15, no. 4, p. 780, 2018.

[4] W. Feng and H. Yuan, "Haze pollution and economic fluctuations: an empirical analysis of Chinese cities," *Cleaner Environmental Systems*, vol. 2, Article ID 100010, 2021.

[5] S. Makgato and E. Chirwa, "Recent developments in reduction of sulphur emissions from selected waterberg coal samples used in south african power plants," *Journal of Cleaner Production*, vol. 276, Article ID 123192, 2020.

[6] M. Yu and Q. Liu, "Deep learning-based downscaling of tropospheric nitrogen dioxide using ground-level and satellite observations," *Science of the Total Environment*, vol. 773, Article ID 145145, 2021.

[7] K. Chen, S. Breitner, K. Wolf et al., "Ambient carbon monoxide and daily mortality: a global time-series study in 337 cities," *The Lancet Planetary Health*, vol. 5, no. 4, pp. e191–e199, 2021.

[8] H. Lu, M. Xie, X. Liu et al., "Adjusting prediction of ozone concentration based on CMAQ model and machine learning methods in Sichuan-Chongqing region, China," *Atmospheric Pollution Research*, vol. 12, no. 6, Article ID 101066, 2021.

[9] P. Sicard, E. Paoletti, E. Agathokleous et al., "Ozone weekend effect in cities: deep insights for urban air pollution control," *Environmental Research*, vol. 191, Article ID 110193, 2020.

[10] A. Mukherjee, D. S. Yadav, S. B. Agrawal, and M. Agrawal, "Ozone a persistent challenge to food security in India: current status and policy implications," *Current Opinion in Environmental Science & Health*, vol. 19, Article ID 100220, 2021.

[11] T. Fischer, "Wheat yield losses in India due to ozone and aerosol pollution and their alleviation: a critical review," *Outlook on Agriculture*, vol. 48, no. 3, pp. 181–189, 2019.

[12] A. L. Power, R. K. Tennant, R. T. Jones et al., "Monitoring impacts of urbanisation and industrialisation on air quality in the anthropocene using urban pond sediments," *Frontiers of Earth Science*, vol. 6, p. 131, 2018.

[13] G. T. H. Nguyen, H. Shimadera, K. Uranishi, T. Matsuo, and A. Kondo, "Numerical assessment of PM2.5 and O3 air quality in Continental Southeast Asia: impacts of potential future climate change2.5 and O3 air quality in continental Southeast Asia: impacts of potential future climate change," *Atmospheric Environment*, vol. 215, Article ID 116901, 2019.

[14] A. M. Fiore, V. Naik, and E. M. Leibensperger, "Air quality and climate connections," *Journal of the Air & Waste Management Association*, vol. 65, no. 6, pp. 645–685, 2015.

[15] G. D'Amato, C. E. Baena-Cagnani, L. Cecchi et al., "Climate change, air pollution and extreme events leading to increasing prevalence of allergic respiratory diseases," *Multidisciplinary Respiratory Medicine*, vol. 8, no. 1, pp. 1–9, 2013.

[16] R. Sivarethinamohan, S. Sujatha, S. Priya, S. Sankaran, A. Gafoor, and Z. Rahman, "Impact of air pollution in health and socio-economic aspects: review on future approach," *Materials Today Proceedings*, vol. 37, pp. 2725–2729, 2021.

[17] P. García Nieto, F. Sánchez Lasheras, E. García-Gonzalo, and F. de Cos Juez, "PM10 concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM, MLP, VARMA and ARIMA: a case study 10 concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM, MLP,

VARMA and ARIMA: a case study," *Science of the Total Environment*, vol. 621, pp. 753–761, 2018.

[18] L. Zhang, J. Lin, R. Qiu et al., "Trend analysis and forecast of PM2.5 in Fuzhou, China using the ARIMA model2.5 in Fuzhou, China using the ARIMA model," *Ecological Indicators*, vol. 95, pp. 702–710, 2018.

[19] A. Kumar and P. Goyal, "Forecasting of air quality in Delhi using principal component regression technique," *Atmospheric Pollution Research*, vol. 2, no. 4, pp. 436–444, 2011.

[20] S. Lwasa, F. Mugagga, B. Wahab, D. Simon, J. Connors, and C. Griffith, "Urban and peri-urban agriculture and forestry: transcending poverty alleviation to climate change mitigation and adaptation," *Urban Climate*, vol. 7, pp. 92–106, 2014.

[21] L. Wu, X. Gao, Y. Xiao, S. Liu, and Y. Yang, "Using Grey Holt-Winters model to predict the air quality index for cities in China," *Natural Hazards*, vol. 88, no. 2, pp. 1003–1012, 2017.

[22] L. Wu, N. Li, and Y. Yang, "Prediction of air quality indicators for the Beijing-Tianjin-Hebei region," *Journal of Cleaner Production*, vol. 196, pp. 682–687, 2018.

[23] W. Yang, J. Wang, T. Niu, and P. Du, "A hybrid forecasting system based on a dual decomposition strategy and multi-objective optimization for electricity price forecasting," *Applied Energy*, vol. 235, pp. 1205–1225, 2019.

[24] W. Yang, J. Wang, H. Lu, T. Niu, and P. Du, "Hybrid wind energy forecasting and analysis system based on divide and conquer scheme: a case study in China," *Journal of Cleaner Production*, vol. 222, pp. 942–959, 2019.

[25] J. Wang, P. Du, Y. Hao, X. Ma, T. Niu, and W. Yang, "An innovative hybrid model based on outlier detection and correction algorithm and heuristic intelligent optimization algorithm for daily air quality index forecasting," *Journal of Environmental Management*, vol. 255, Article ID 109855, 2020.

[26] W. j. Niu and Z. k. Feng, "Evaluating the performances of several artificial intelligence methods in forecasting daily streamflow time series for sustainable water resources management," *Sustainable Cities and Society*, vol. 64, Article ID 102562, 2021.

[27] N. Singh, S. Sugandha, T. Mathur, S. Agarwal, and K. Tiwari, "Stock price prediction using fractional gradient-based long short term memory," in *Journal of Physics: Conference Series*vol. 1969, IOP Publishing, Article ID 12038, 2021.

[28] R. Janarthanan, P. Partheeban, K. Somasundaram, and P. Navin Elamparithi, "A deep learning approach for prediction of air quality index in a metropolitan city," *Sustainable Cities and Society*, vol. 67, Article ID 102720, 2021.

[29] K. S. Pritam, S. Sugandha, T. Mathur, and S. Agarwal, "Underlying dynamics of crime transmission with memory," *Chaos, Solitons & Fractals*, vol. 146, Article ID 110838, 2021.

[30] K. Bansal, S. Arora, K. S. Pritam, T. Mathur, and S. Agarwal, "Dynamics of Crime Transmission Using Fractional Order Differential Equations," *Fractals*, vol. 30, no. 1, pp. 2250012–1485, 2022.

[31] I. Podlubny, "Fractional-order systems and fractional-order controllers," *Institute of Experimental Physics, Slovak Academy of Sciences, Kosice*, vol. 12, no. 3, pp. 1–18, 1994.

[32] M. Caputo, "Linear models of dissipation whose Q is almost frequency independent—II," *Geophysical Journal International*, vol. 13, no. 5, pp. 529–539, 1967.

[33] Y. A. Rossikhin and M. Shitikova, "Fractional calculus models in dynamic problems of Viscoelasticity," *Handbook of Fractional Calculus with Applications*, vol. 7, pp. 139–158, 2019.

[34] C. A. Valentim Jr, N. A. Oliveira, J. A. Rabi, and S. A. David, "Can fractional calculus help improve tumor growth models?" *Journal of Computational and Applied Mathematics*, vol. 379, Article ID 112964, 2020.

[35] S. Arora, T. Mathur, S. Agarwal, K. Tiwari, and P. Gupta, "Applications of fractional calculus in computer vision: a survey," *Neurocomputing*, vol. 489, pp. 407–428, 2022.

[36] H. Wang, "Research on application of fractional calculus in signal real-time analysis and processing in stock financial market," *Chaos, Solitons & Fractals*, vol. 128, pp. 92–97, 2019.

[37] V. E. Tarasov, "On history of mathematical economics: application of fractional calculus," *Mathematics*, vol. 7, no. 6, p. 509, 2019.

[38] H. Sun, Y. Zhang, D. Baleanu, W. Chen, and Y. Chen, "A new collection of real world applications of fractional calculus in science and engineering," *Communications in Nonlinear Science and Numerical Simulation*, vol. 64, pp. 213–231, 2018.

[39] L. Wu, S. Liu, L. Yao, S. Yan, and D. Liu, "Grey system model with the fractional order accumulation," *Communications in Nonlinear Science and Numerical Simulation*, vol. 18, no. 7, pp. 1775–1785, 2013.

[40] S. Metia, S. D. Oduro, H. N. Duc, and Q. Ha, "Inverse air-pollutant emission and prediction using extended fractional Kalman filtering," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 2051–2063, 2016.

[41] Y. Sun, X. Wu, J. Cao, Z. Wei, and G. Sun, "Fractional extended Kalman filtering for non-linear fractional system with lévy noises," *IET Control Theory & Applications*, vol. 11, no. 3, pp. 349–358, 2017.

[42] Y. V. Oktaviana, E. Apriliani, and D. K. Arif, "Fractional Kalman filter to estimate the concentration of air pollution," in *Journal of Physics: Conference Series*vol. 1008, IOP Publishing, Article ID 12008, 2018.

[43] Y. F. Pu, Z. Yi, and J. L. Zhou, "Fractional hopfield neural networks: fractional dynamic associative recurrent neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2319–2333, 2017.

[44] J. Wang, Y. Wen, Y. Gou, Z. Ye, and H. Chen, "Fractional-order gradient descent learning of BP neural networks with Caputo derivative," *Neural Networks*, vol. 89, pp. 19–30, 2017.

[45] H. Wang, Y. Yu, G. Wen, S. Zhang, and J. Yu, "Global stability analysis of fractional-order hopfield neural networks with time delay," *Neurocomputing*, vol. 154, pp. 15–23, 2015.

[46] S. Khan, J. Ahmad, I. Naseem, and M. Moinuddin, "A novel fractional gradient-based learning algorithm for recurrent neural networks," *Circuits, Systems, and Signal Processing*, vol. 37, no. 2, pp. 593–612, 2018.

[47] X. Chen, "Application of Fractional Calculus in Back Propagation Neural Networks," Ph.D Thesis, Nanjing Forestry University, Nanjing, China, 2013.

[48] S. Khan, I. Naseem, M. A. Malik, R. Togneri, and M. Bennamoun, "A fractional gradient descent-based RBF neural network," *Circuits, Systems, and Signal Processing*, vol. 37, no. 12, pp. 5311–5332, 2018.

[49] D. Sheng, Y. Wei, Y. Chen, and Y. Wang, "Convolutional neural networks with fractional order gradient method," *Neurocomputing*, vol. 408, pp. 42–50, 2020.

[50] G. Jumarie, "Modified Riemann-Liouville derivative and fractional Taylor series of nondifferentiable functions further results," *Computers & Mathematics with Applications*, vol. 51, no. 9-10, pp. 1367–1376, 2006.

[51] G. Jumarie, "Table of some basic fractional calculus formulae derived from a Modified Riemann-Liouville derivative for non-differentiable functions," *Applied Mathematics Letters*, vol. 22, no. 3, pp. 378–385, 2009.

[52] G. Jumarie, "An approach via fractional analysis to non-linearity induced by coarse-graining in space," *Nonlinear Analysis: Real World Applications*, vol. 11, no. 1, pp. 535–546, 2010.

[53] G. Jumarie, "On the derivative chain-rules in fractional calculus via fractional difference and their application to systems modelling," *Open Physics*, vol. 11, no. 6, pp. 617–633, 2013.

[54] Y. J. Cruz, M. Rivas, R. Quiza, R. E. Haber, F. Castaño, and A. Villalonga, "A two-step machine learning approach for dynamic model selection: a case study on a micro milling process," *Computers in Industry*, vol. 143, Article ID 103764, 2022.