Sigma Journal of Engineering and Natural Sciences, Vol. X, No. X, pp. XX-XX, March, 2022
Yildiz Technical University Press, Istanbul, Turkey

Role of air pollution on COVID-19 in Istanbul

İlknur Dönmez[1] [0000-0002-8344-1180] Zafer Aslan[2] [0000-0001-7707-7370]

[1] İstanbul Arel University, Faculty of Engineering and Architecture, Computer Engineering Department, 34537, Tepekent, İstanbul; zaferaslan@aydin.edu.tr
[2] İstanbul Aydın University, Faculty of Engineering, Computer Engineering Department, 34295, Florya, İstanbul; ilknurdonmez@arel.edu.tr

ABSTRACT

There are many types of research on strong relations between air pollution and the respiration system. This paper is related to the effect of air pollution on pandemic case numbers. By analyses of the number of patients and mean air pollution ($\mu g/m^3$) data by using data mining, it is concluded that there was evidence of this relationship with a significant level alpha = 0,10. This paper estimates "the number of patients with coronavirus" as a function of daily air pollution and corona case numbers data using single ANN, LSTM, and W-ANN, W-LSTM hybrid methods. This paper explains small, meso, and large scale factors and their role on COVID-19 patients. This finding does not demonstrate a direct cause-effect relationship between air pollution and COVID-19 patients. This study shows the importance of pollution on the number of patients with COVID -19 infection. Although our results have significant uncertainties, we can clearly distinguish the contribution of air pollution to COVID-19 patients. When we used wavelet transformation of air pollution data to estimate COVID-19 patient's numbers, R^2 score is increased in both ANN and LSTM between [0.01- 0.10]. Nevertheless, the actual number of patients is influenced by many additional factors such as the country's health system. After error analysis, sMAPE [3.6-5.9] changed, there is sufficient evidence of model results (M3, M4 Hybrid LSTM) and observation. ($0.91<R^2<0.96$, alpha=<0.01). The performance of hybrid model is 10% better than the simple ANN model.

Keywords: *Air pollution, COVID-19, ANN, LSTM, Wavelet.*

INTRODUCTION

According to Thomas Münzel (2020), when people inhale polluted air, the very small polluting particles migrate from the lungs to the blood and blood vessels causing inflammation and severe oxidative stress [1]. Like air pollution, COVID-19 damages the lining of the arteries, endothelium, and causes the arteries to narrow and harden. In case both long-term exposure to air pollution and COVID-19 infection come together, this harms wellbeing, especially concerning the heart and blood vessels, which leads to more defenselessness and less resilience to COVID-19. If people have heart disease, both air pollution, and coronavirus infection will increase disorders that can lead to heart attacks, heart failure and, strokes. Air pollution damages the lungs and increases the activity of ACE-2, which allows the virus to travel more easily through the lungs and possibly blood vessels and heart. Poor air quality is one of the leading risk factors for many diseases, especially due to fine particulate matter <2.5 mm in diameter ($PM_{2.5}$) [2,3]. The loss in global average life expectancy rates due to prolonged exposure to ambient air pollution exceeds the infectious disease impact and reaches levels comparable to tobacco smoking [4].

A study in 2003 [5] analyze the first severe acute respiratory syndrome coronavirus (SARS-CoV-1) outcomes. They found that the risk of dying from the disease in some parts of China with medium-level air pollution was 80% higher than in areas with relatively fresh air. In China, the coronavirus (SARS-CoV-2) was named

This paper was recommended for publication in revised form by Regional Editor Sania Qureshi
[1] Department of Computer Engineering, Arel University, Istanbul, Turkey
[2] Department of Computer Engineering, Aydın University, Istanbul, Turkey
**E-mail address: ilknurdonmez@arel.edu.tr, zaferaslan@aydin.edu.tr*
Orcid id: https://orcid.org/0000-0002-8344-1180; https://orcid.org/0000-0001-7707-7370;
Manuscript Received 10 May 2021, Accepted 12 January 2022

COVID19 in 2019, and it turned from an epidemic to a pandemic in early 2020. COVID-19 is related to a combination of respiratory and cardiovascular complications, including increases in biomarkers [6] due to myocardial infarction, heart failure, venous thromboembolism. These are also found in exposure to high levels of air pollutants [7].

Considering the effects of air pollution on cardiovascular and respiratory health, the association with COVID-19 is not unexpected. Many studies have addressed the impact of air pollution on COVID19 in different regions. In China, the incidence of COVID-19 was found to be significantly enhanced by $PM_{2.5}$ [8], while a correlation between ambient $PM_{2.5}$ and the mortality rate was also established [9]. In Italy, it was found that the high pollution concentrations that are typical for the Povalley in the Lombardy region of which Milan is the capital, were associated with high case numbers and mortality rate [10]. In the USA the severity of COVID-19 outcomes was linked to $PM_{2.5}$ exposures, making use of Medicare data for >60 million people and nationwide air quality measurements [11]. In heavily polluted areas, the risk was twice as high compared with areas having relatively fresh air.

A new study using an ecological design evaluated how environmental impacts are altering the severity of COVID-19 consequences in the United States. In this study, researchers explored whether long-term mean exposure to fine particulate matter ($PM_{2.5}$) is related to illness and an expanded risk of COVID-19 death rates in the USA. They found that an increment of 1 g/m3 in $PM_{2.5}$ was related to an 8% increment in COVID-19 death rate with 95% confidence interval [CI]: 2%, 15% which is statistically significant and robust to secondary and sensitivity analyses. In a study [12] a dynamic model is applied to predict the association between air quality and COVID-19 cases. ANN has been applied to air pollution prediction by lockdown level. Because of the limited database, they prove that the Multilayer Perceptron neural network is robust with a Mean Absolute Percentage Error ~ 30%. In another research [13], French cities (Paris, Lyon, and Marseille) air pollution and corona numbers are analyzed to investigate the relationship between the Coronavirus Disease 19 (COVID-19) outbreak and air pollution. They used Artificial Neural Networks (ANNs) to determine the concentration of $PM_{2.5}$ and PM_{10} linked to COVID-19-related deaths. They perform a D2C (Causal Direction from Dependency) algorithm capable of predicting the existence of a direct causal link between two variables in a multivariate setting. In a study in China, the researcher collected the daily COVID-19 death number, air quality index (AQI), ambient air pollutant concentrations, and meteorological variables data of Wuhan between Jan 25 and April 7, 2020. They used Pearson and Poisson regression models to understand the association between COVID-19 deaths and each risk factor [14].

In another research, four cities in Italy were selected as a case study, and some notable climate parameters such as daily average temperature, relative humidity, wind speed, and an urban parameter, population density, were considered as input data set. The ANN, PSO, and DE algorithms are applied to select the best parameters to reach output which is confirmed cases of COVID-19 [15]. In another study in Germany using daily data from February 24, 2020, to July 02, 2020, the researchers found that $PM_{2.5}$, O_3, and NO_2 have a significant relationship with the outbreak of COVID-19 case numbers and deaths [16]. The Gaussian approach is used for probability and correlation between the number of COVID-19 cases and the air pollution in Lima, Peru [17]. Researchers claim that the correlation between NO_2 and infections and PM_{10} has been described with R: 98.827% and 95.38% in this region. In the last study [18] the researcher investigated whether long-term exposure to air pollution increases the risk of COVID-19 infection in Germany. They found that nitrogen dioxide (NO_2) is significantly associated with COVID-19 incidence, with a 1 g/m^3 increase in long-term exposure to NO_2 increasing the COVID-19 incidence rate by 5.58% (95% credible interval [CI]: 3.35%, 7.86%).

Particulate matter (PM), also called particle pollution, is a general term for extremely small particles and liquid droplets in the atmosphere. $PM_{2.5}$ are inhalable particles less than 2.5 microns in diameter and PM_{10} are fine inhalable particles less than 10 microns in diameter. Even air pollution can take lots of different forms, like carbon compounds such as carbon monoxide (CO), carbon dioxide (CO_2), sulfuric compounds like sulfur dioxide (SO_2), methane, radioactive decay, or toxic chemicals, there are lots of studies that propose Particulate matter pollution as a major cause of air quality illnesses [19,20,21]. There are also lots of studies on 2021 based on the mathematical model of coronavirus, air pollution, dust [22,23] and smoke [24,25] and transmission of coronavirus. In a study, a model that contains six nonlinear fractional-order differential equations is used [26]. The objective of the study is

optimal control of the model they proposed. In two different studies, the mathematical model of the transmission dynamics of COVID-19 is analyzed [27,28,29]. The fractional COVID-19 epidemic model is analyzed under the Caputo Operator by Rahat Zarin et al. [30] and a fuzzy-based strategy to suppress the novel coronavirus (2019-ncov) massive outbreak is proposed by Hadi Jahanshahi et al [31]. In studies, Dynamical Aspects of Smoking Model is analyzed mathematically [32,33]. There are also some studies that analyzes the air quality parameters related with COVID-19 [34,35].

In our study, we try to analyze the relation between COVID-19 case numbers and mean PM values of air pollution data using synchronized daily data for 3 months for İstanbul.

Istanbul, which is a city of history and culture, has also the largest population in Europe. In addition to industry, Istanbul, where it is concentrated in commercial activities, acts as a center between North, South, East and West due to its location. Istanbul is the most populous city in Turkey with a population of around 16 million. The high number of human activities that come with the excess of the population also directly affect the PM10 concentration, which is one of the pollutants that cause air pollution. In our study as machine learning algorithms, two different deep learning method and wavelet hybrid method is used to analyze and predict the COVID-19 case numbers using pollution data. We proposed that COVID-19 case numbers are affected by air pollution. So using ANN and LSTM auto-regression method and their hybrid with Wavelet, we try to predict case numbers, using previous case number data.

The novelty of this study is using wavelet and LSTM together to predict corona numbers using air pollution data. As far as we have researched, there is no similar study in the literature. The motivation behind the model is that wavelet transformed data gives better accuracy in time series predictions in different type of data so we used it on pollution data for COVID-19 case numbers prediction. When we also used synchronized air pollution data as input, our case number predictions accuracy was [0-0,10] higher.

In Section-2, the data and the used ANN, LSTM, Wavelet, and hybrid methods are explained. Section-3 is related to the results and the evaluation of each model. The conclusion part is in Section-4.

MATERIAL AND METHODS

The data and methods used in this study are described below.

STUDY AREA AND DATA WAVELET THEORY AND APPLICATION

In this study, İstanbul which is the most crowded city in Turkey is taken into account. İstanbul daily air pollution data and daily corona case numbers in İstanbul for 3 months are processed. To get reliable and valid data we used the shared corona data by the Turkish Ministry of Health [36]. The air pollution data is taken from Municipality of Istanbul [37]. The daily average of pollution data from 29.06.2020 to 30.09.2020 was taken into account. Minimum value of mean particle matter pollution is 89,3208 g/m^3 and maximum value of mean particle matter pollution is 326.4481 g/m^3 in this time period.

WAVELET THEORY AND APPLICATION

Most signals are represented in the time domain. More information about the time signals can be obtained by applying signal analysis. The Fourier transform is the most commonly known method to analyze a time signal for its frequency content. A relatively new analysis method is wavelet analysis. The wavelet analysis differs from the Fourier analysis by using short wavelets instead of long waves for the analysis function [38]. The wavelet analysis has some major advantages over Fourier transform which makes it an interesting alternative for many applications. The Fourier transform does not give satisfactory results for signals that are highly non-stationary, noisy, a-periodic, etc. These types of signals can be analyzed using local analysis methods. These methods include the short-time Fourier transform and the wavelet analysis. The analysis of a non-stationary signal using the FT or the STFT does not give satisfactory results. Better results can be obtained using wavelet analysis. One advantage of wavelet analysis is the ability to perform local analysis and to present results at frequency and time domain on the same graph[39]. Wavelet analysis can reveal signal aspects that other analysis techniques miss, such as trends, breakdown points, discontinuities, extremes etc. In comparison to the STFT, wavelet analysis makes it possible to perform a multi-resolution analysis. The use and fields of application of wavelet analysis have grown rapidly in the last years [40,41].

In Discrete Wavelet Transform, the DWT of a signal "x" is calculated by passing it through a series of filters. As seen on the Eq1, first, the samples are passed through a low pass filter with impulse response "f" resulting in a convolution of the two:

$$y(n) = (x * f)(n) == \sum_{k=-\infty}^{\infty} x(k)f(n-k) \quad (1)$$

where y(n) is DWT of signal x, and n is the level of decomposition.

The signal is also decomposed simultaneously using a high-pass filter h. The outputs give the detail coefficients (from the high-pass filter) and approximation coefficients (from the low-pass). It is important that the two filters are related to each other and they are known as quadrature mirror filters.

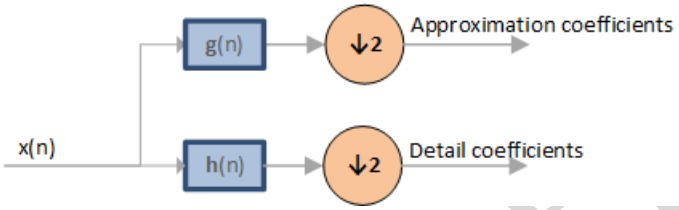

Fig. 1 Wavelet inner structure.

However, since half the frequencies of the signal have now been removed, half the samples can be discarded according to Nyquist's rule. The filter output of the low-pass filter g in the Fig1 is subsampled by 2 and further processed by passing it again through a new low-pass filter "g" and a high-pass filter h with half the cut-off frequency of the previous one as seen in the Eq2 and Eq3.

$$y_{low}(n) = \sum_{k=-\infty}^{\infty} x(k)g(2n-k) \quad (2)$$

$$y_{high}(n) = \sum_{k=-\infty}^{\infty} x(k)g(2n-k) \quad (3)$$

where $y_{low}(n)$ is low pass filter output and $y_{high}(n)$ is high pass filter output..

ARTIFICIAL NEURAL-NETWORK (ANN) THEORY AND APPLICATION

ANNs have been promising applications in different engineering fields [42,43]. It is still used in lots of recent studies [44,45]. In the research, the ANN model was implemented with Python Keras packages using the feed-forward back propagation (BP) method. The model is trained using the input generated from air pollution data and previous case numbers data, and the output generated from corona case numbers.

In this study, a total of 90 samples were divided into 2 datasets namely, training, and testing. Data were randomly partitioned with 80 % (72 samples) for training, and 20 % (18 samples) for the testing dataset. After the training, the ANN was validated using the test data to evaluate the prediction performance of the final model.

In the study, hidden neurons in the range of 2–24 were taken for analysis based on hit and trial. The neurons were optimized to get the best prediction performance. Throughout the training process, the connection weights and biases were adjusted to minimize the difference between the experimental and predicted values. The neurons of the hidden and output layer act as a summing junction that joins and adjusts the inputs from the preceding layer using Eq4, Eq5 and Eq6,

$$H_{1,j} = \sum_{i=1}^{4} X_i w_{1,i,j} \quad (4)$$

$$H_{2,k} = \sum_{j=1}^{12} H_{1,j}\, w_{2,j,k} \quad (5)$$

$$y = \sum_{k=1}^{12} H_{2,k} w_{3,k} \quad (6)$$

where j $H_{1,j}$ is the jth node of the first hidden layer, $H_{2,k}$ is the kth node of the second hidden layer and $w_{1,i,j}$ is the weight between X_i and $H_{1,j}$; $H_{2,k}$ is the jth node of the last hidden layer, $w_{2,j,k}$ is the weight between the hidden layer elements. $w_{3,k}$ is the weight between the second hidden layer and output. The developed neural-network model utilizes a sigmoidal transfer function to measure the nonlinear relationship at the output.

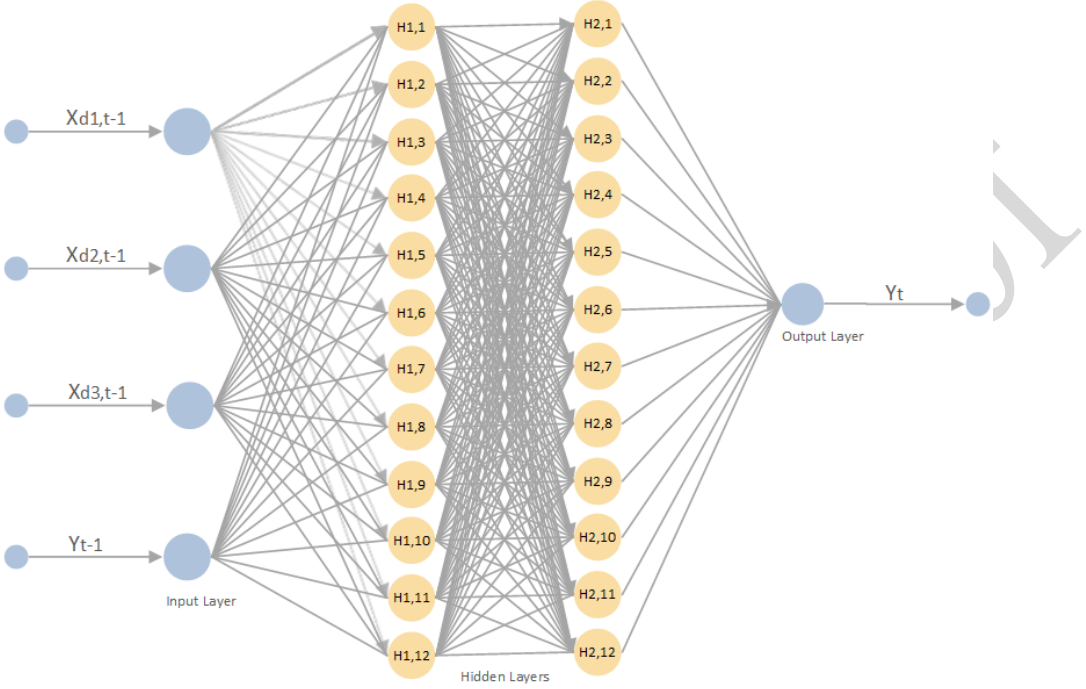

Fig. 2 ANN topology.

In our python application, we used the topology of one input layer, two hidden layers, and one output layer (1x12x12x1) as the best according to accuracy for ANN structure as seen in the Fig2. In inner layers, the Relu activation function for the output Sigmoid activation function is used. And we used the mean square error for lost function, Adam as an optimizer, and 800 iterations for the 4-layer ANN model.

LONG SHORT-TERM MEMORY (LSTM) THEORY AND APPLICATION

LSTM is a Recurrent Artificial Neural Network (RANN) that has been widely used in the deep learning field [46]. The main advantage of the LSTM over conventional feed-forward neural networks is its capability to remember patterns over a long time because of its advanced structure which depends on feedback connections [47,48]. It also can alleviate the problem of vanishing gradient which is a critical issue in other RANN models. LSTM networks have the general capability to deal with sequence prediction and forecasting problems. The architecture of the LSTM network, shown in Fig2, is an improved version of a RANN architecture which is trained using the back-propagation technique [46]. LSTM uses multiple gates that act as memory cells to manage the flow of the inputs and outputs inside the network instead of a conventional hidden layer architecture with non-linear activation functions. Thus, LSTM has a complex hidden layer that contains many parameters compared with conventional RANN.

In RNN previous cell output h_{t-1} and input x_t goes into the tanh function to generate new cell output h_t. While the traditional RNN cell has two inputs previous cell output and current input, the LSTM cell has three. Each cell of the LSTM has three internal layers. As seen in Fig3 and Fig4, the first section of the cell is the "forget gate" that controls what information is maintained from the previous state. This takes in the previous cell output h_{t-1} and the current input x_t band and applies a sigmoid activation layer ( ) to get values between 0 and 1 for each hidden unit. This is followed by element-wise multiplication with the current state.

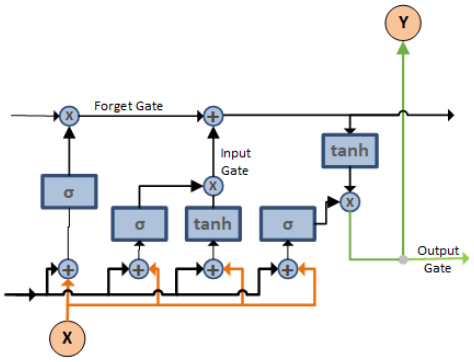

Fig3 LSTM cell structure.

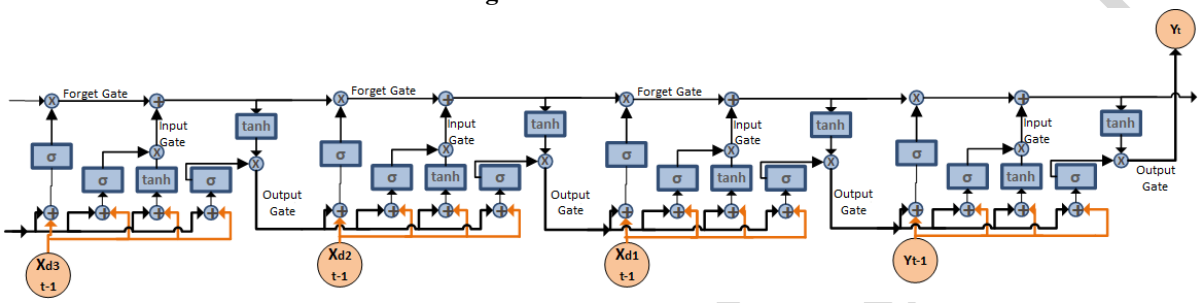

Fig4 LSTM topology.

The next layer is an "input gate" that updates the state based on the current input. This passes the same input (h_{t-1} and x_t) into a sigmoid activation layer (σ) and a tanh activation layer (tanh) and performs element-wise multiplication between these two results. Next, element-wise addition is performed with the result and the current state after applying the "forget gate" to update the state with new information. Finally, we have an "output gate" that controls what information gets passed to the next state. We run the current state through a tanh activation layer (tanh) and perform element-wise multiplication with the cell input (h_{t-1} and x_t) run through a sigmoid layer (σ) that acts as a filter on what we decide to output. This output y_t is then passed to the LSTM cell for the next input of our sequence and also passed up to the next layer of our network.

HYBRID W-ANN, W-LSTM THEORY AND APPLICATION

As hybrid W-ANN and W-LSTM method, first, we found the wavelet components of the input time-series data using Discrete Wavelet Transformation. Then we used this component as the input of the classical W-ANN and W-LSTM method.

EVALUATION AND RESULTS

In this section Wavelet analysis of pollution concentration and COVID-19 patient numbers results, ANN, LSTM, and hybrid methods results are discussed and explained.

WAVELET ANALYSIS RESULTS

Wavelet analysis of pollution concentration: In Fig5a, for analyzing the total pollution concentration we used wavelet decomposition of air pollution data at level 3. The data (s) is composed of a_3, d_1, d_2 and d_3. From the original signal, a gradually increasing trend is observed in pollution in the study area. Large scale factors are low at the middle part of the study term. The role of local-scale factors has been observed in all periods. Small scale factors and their role is lower than another period at the beginning of the second half of the study period.

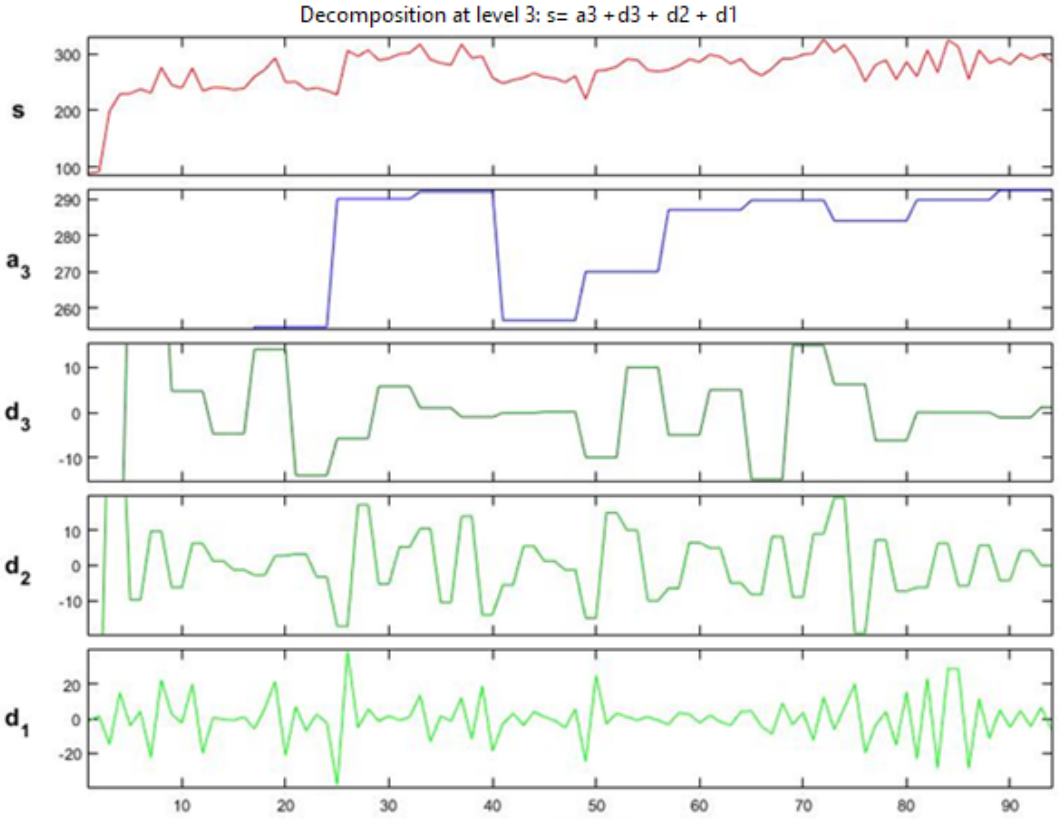

Fig5a. Temporal variation of pollution mean air pollution (μg/m³), from 29.06.2020 to 30.09.2020 1D Wavelet, Db.

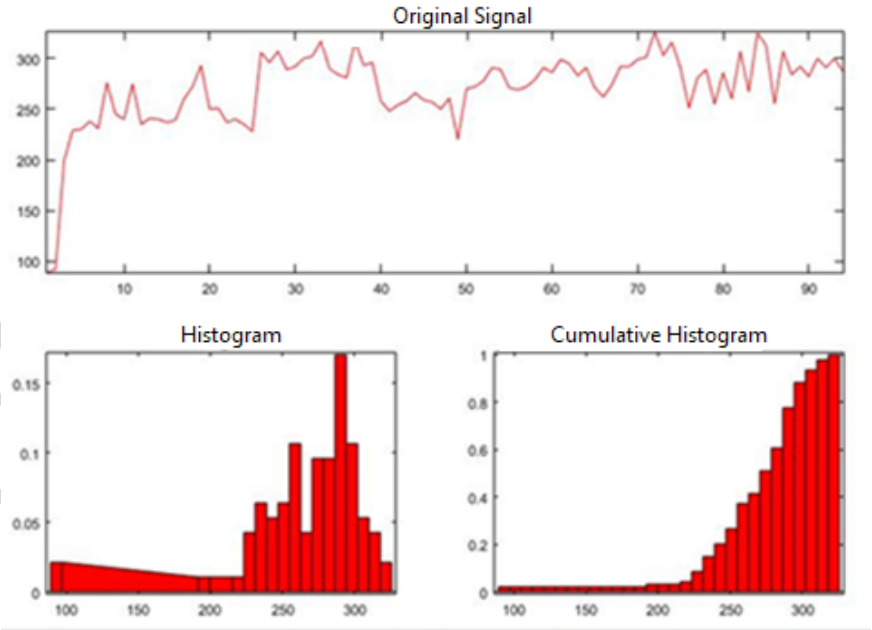

Fig5b. Descriptive statistics of mean air pollution, from 29.06.2020 to 30.09.2020 1D

Fig5b shows histogram and descriptive statistics of air pollution in the study area. There is a negative skewness. Mean, maximum, minimum, and median values are 270.2, 326.4, 89.3, and 277 respectively.

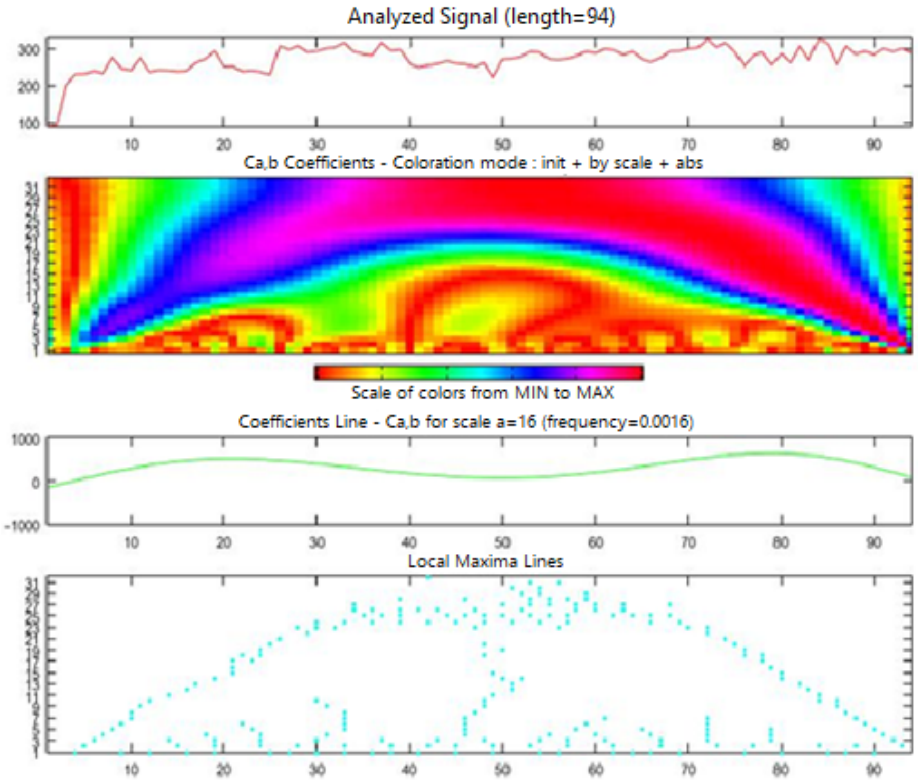

Fig5c. 1D Continuous wavelet, Mexh, sampling period 1, mean air pollution (g/m³), from 29.06.2020 to 30.09.2020

Temporal variation of air period at all periods is under the effects of large, meso, and small scale factors. 5 - 14 days periodicity is available on temporal variation of pollution concentration as seen in Fig5c. The minimum value of concentration is associated with the effects of local and micro-scale factors. (in the middle part of the study period).

Wavelet analysis of patient numbers: In Fig6a, for analyzing the pollution concentration we used wavelet decomposition of air pollution data at level 3. The data (s) is composed of a3, d1, d2, and d3. From the original signal, a gradually increasing trend is observed in pollution in the study area.

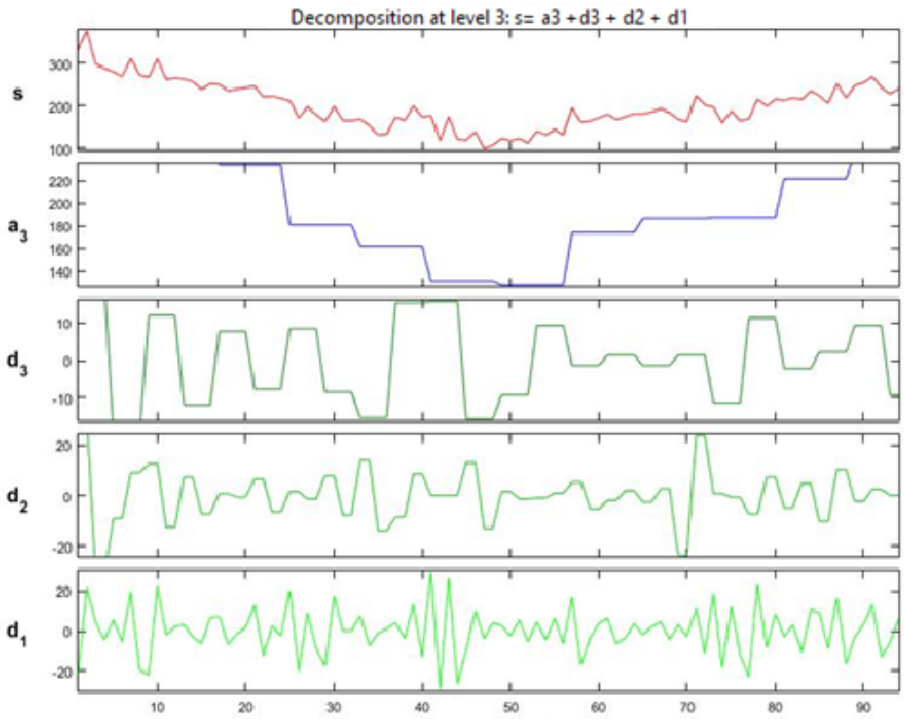

Fig6a. Number of Patient based on COVID-19 virus, 1D wavelet db, from 29.06.2020 to 30.09.2020 1D.

Fig6a shows the variation of patient numbers. All periods there are combined effects of small, meso, and large scale factors on the number of patients. The minimum patient number at the middle part of the period is mainly associated with small and large scale factors. Maximum numbers of the patient are under the effects of all three scale factors.

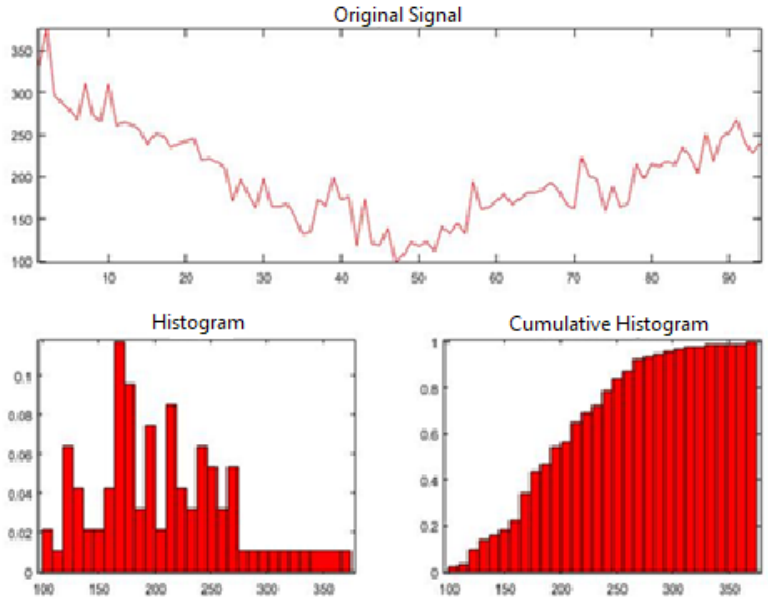

Fig6b. Descriptive statistics of patients' number.

As seen in Fig6b, min, max, median and mean values of patients are 99, 375, 196.5, 200.8 respectively. Positive skewness and multimodal variations have been observed as seen in Fig6b.

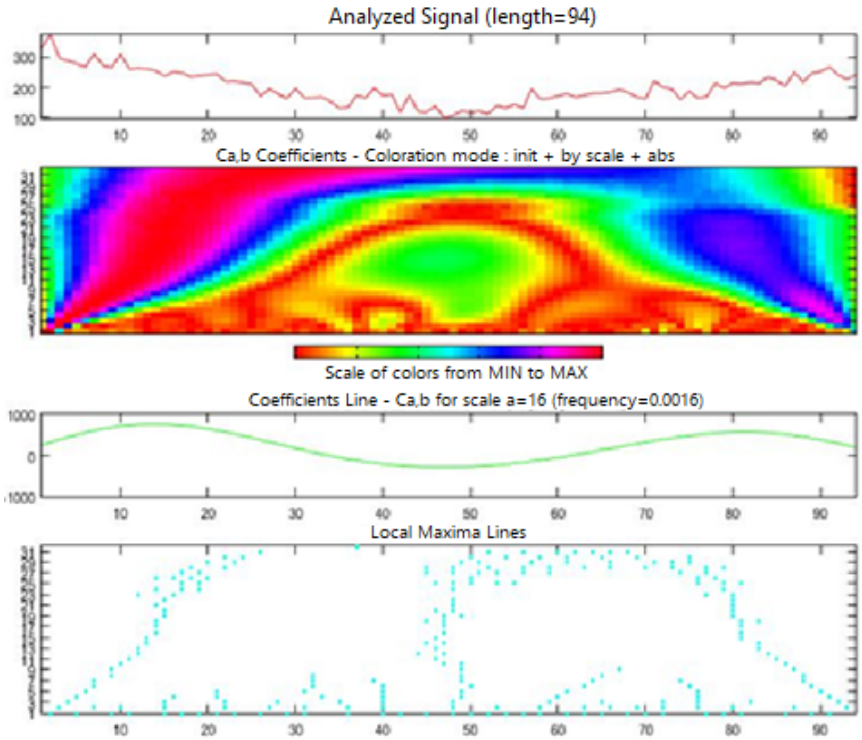

Fig6c. 1D Analyses of patient numbers by Continuous Wavelet. Mexh, sampling 1.

Maximum values are associated with large-scale factors (at the beginning and the end of the period, blue and purple). Small and mesoscale factors (With the periodicity with around 3-20 days, have a role on the minimum value in the middle part of the period, green area) as seen in the Fig6c.

Table 1. Mean values of wavelet details (d1, d2, d3, small, meso and large scale factors) for pollution and patient numbers

Variable	d1	d2	d3
Mean Air Pollution	0	2.34^-17	-0.024
Number of Patient	0	-1.310^-16	0.197

Table 1 shows the role of small, meso, and large scales and energy transfer at all levels. Main energy is associated with small-scale factors (with the periodicity around) at two variables. The weighting values of each detail are shown in Table 2.

Table 2. Relation with variables and details (small, meso and large scale factors)

Variable/correlation coefficient r^2	d1	d2	d3
Mean Air Pollution	0.124	0.117	0.167
Number of Patient	0.048	0.036	0.04875

The number of patients is related to all details. But there is not significant relation. Pollution concentration is related with all details with alpha, “0.10”. Higher relations between air pollution and number of patient are defined between large, small and meso scale factors respectively.

COVID-19 CASE NUMBER PREDICTIONS WITHOUT USING AIR POLLUTION DATA

We used R-squared (R^2) score and symmetric mean absolute percentage error (sMAPE) value [49] to evaluate the models. R^2 is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable. R-squared explains to what extent the variance of one variable explains the variance of the second variable [50,51]. sMAPE is an accuracy measure based on percentage (or relative) errors. The sMAPE is calculated using Eq.7 where R_t is the real value and P_t is the predicted value.

$$sMAPE = \frac{100\%}{n} + \sum_{t=1}^{n} \frac{|P_t - R_t|}{(|R_t| + |P_t|)/2} \quad (7)$$

In our first model (M.1), we predict the current COVID-19 case number with previous COVID-19 case numbers using ANN and LSTM. In this autoregressive model y_t is a function of the lags of y_t where, y_{t-1}, y_{t-2} are the lag1 and lag2 of the time series data, as seen in Eq.8. ANN and LSTM prediction results can be seen in Fig7a and Fig7b respectively.

$$y_t = f(y_{t-1}, y_{t-2}, c) \quad (8)$$

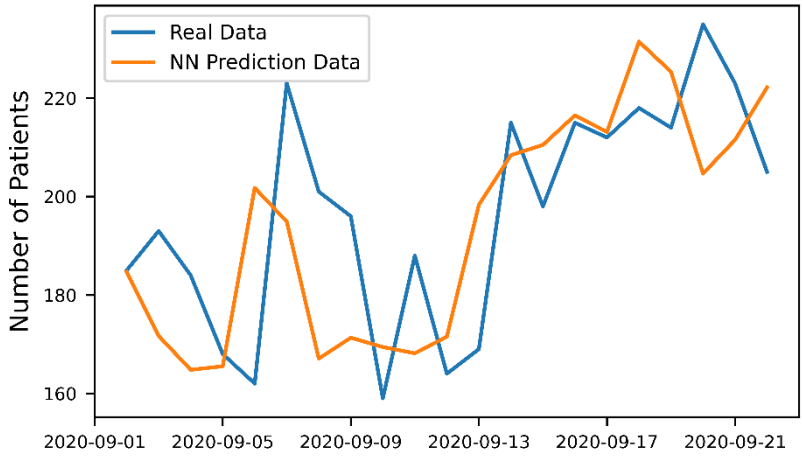

Fig7a ANN results of $f(y_{t-1}, y_{t-2}, c)$.

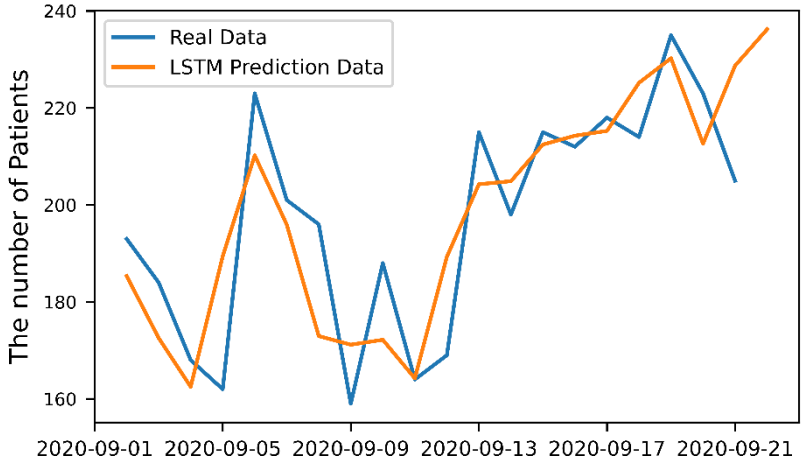

Fig7b LSTM results of $f(y_{t-1}, y_{t-2}, c)$.

COVID-19 CASE NUMBER PREDICTIONS USING AIR POLLUTION DATA

In our second model (M.2), we predict the current COVID-19 case number with previous COVID-19 case numbers and air pollution data using ANN and LSTM. In this model, y_t is a function of the lags of y_t and also x_{t-1}

where, y_{t-1}, y_{t-2} are the lag1 and lag2 of the COVID-19 time series data and x_{t-1}, is the lag for the air pollution time series data as seen in Eq.9. ANN and LSTM prediction results can be seen in Fig8a and Fig8b respectively.

$$y_t = f(x_{t-1}, y_{t-1}, y_{t-2}, c) \quad (9)$$

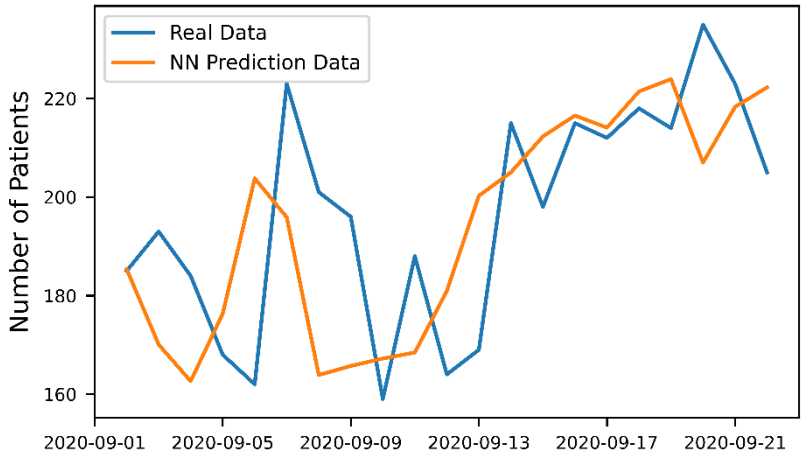

Fig8a ANN results of $f(x_{t-1}, y_{t-1}, y_{t-2}, c)$.

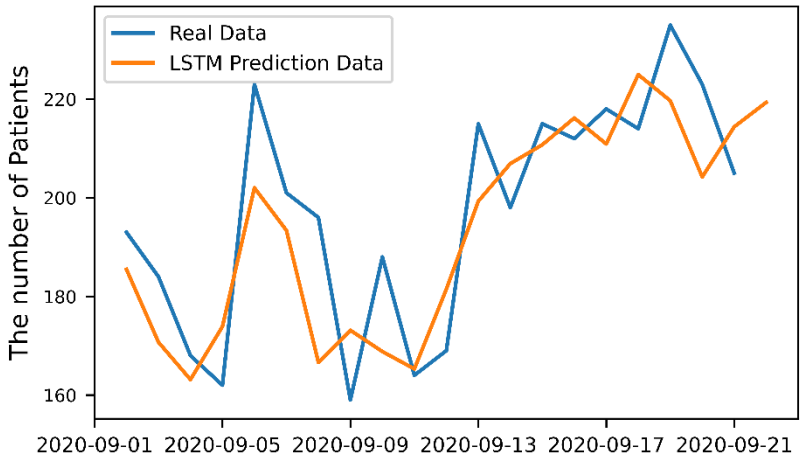

Fig8b LSTM results of $f(x_{t-1}, y_{t-1}, y_{t-2}, c)$.

COVID-19 CASE NUMBER PREDICTIONS WITH USING HYBRID METHODS

In our third model (M.3), we predict the current COVID-19 case number with previous COVID-19 case numbers and wavelet transformed air pollution data using ANN and LSTM. In this model y_t is a function of the lags of y_t (y_{t-1}) and also three wavelet transformation of x_{t-1}, which are $x_{d1,t-1}, x_{d2,t-1}, x_{d3,t-1}$ as seen in Eq.10. Here y_t is time-dependent data of COVID-19 time series, x_t is time-dependent data of air pollution time series. ANN and LSTM prediction results can be seen in Fig9a and Fig9b respectively.

$$y_t = f(x_{d1,t-1}, x_{d2,t-1}, x_{d3,t-1}, y_{t-1}, c) \quad (10)$$

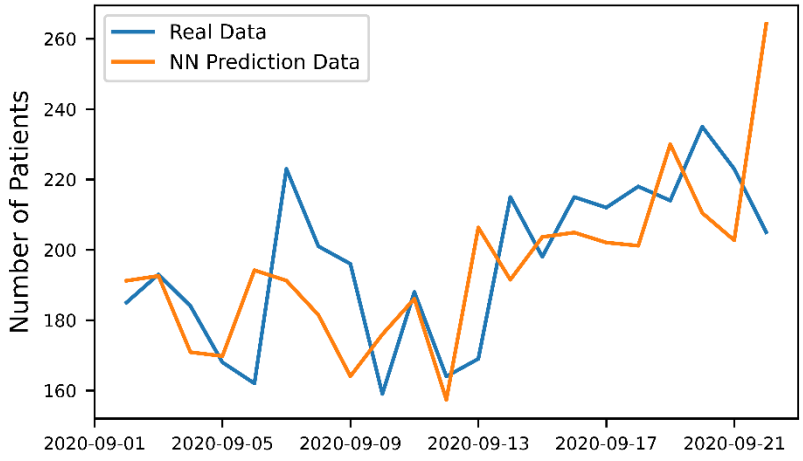

Fig9a Hybrid ANN results of $f\left(x_{d1,t-1}, x_{d2,t-1}, x_{d3,t-1}, y_{t-1}, c\right)$.

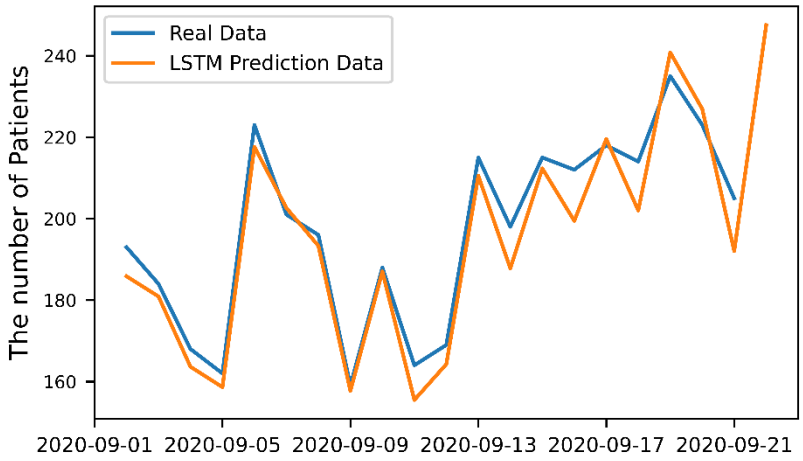

Fig9b Hybrid LSTM results of $f\left(x_{d1,t-1}, x_{d2,t-1}, x_{d3,t-1}, y_{t-1}, c\right)$.

COVID-19 CASE NUMBER PREDICTIONS WITH USING HYBRID METHODS

As a result of our wavelet transformation analysis, $x_{d2,t-1}$, which is one of the three wavelet transformations of x_{t-1} is removed. Its effect to the R^2 score is searched. In the last model (M.4), y_t is a function of $x_{d1,t-1}, x_{d3,t-1}, y_{t-1}$ as seen in Eq.11. ANN and LSTM prediction results can be seen in Fig10a and Fig10b respectively.

$$y_t = f\left(x_{d1,t-1}, x_{d3,t-1}, y_{t-1}, c\right) \quad (11)$$

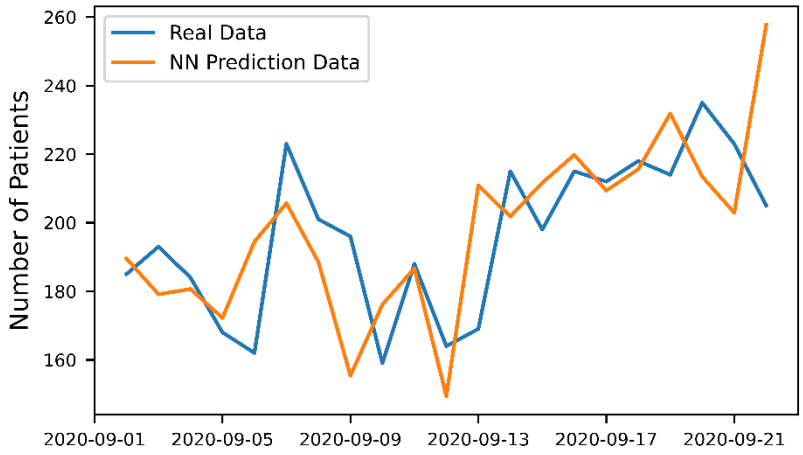

Fig10a Hybrid ANN results of $f(x_{d1,t-1}, x_{d3,t-1}, y_{t-1}, c)$.

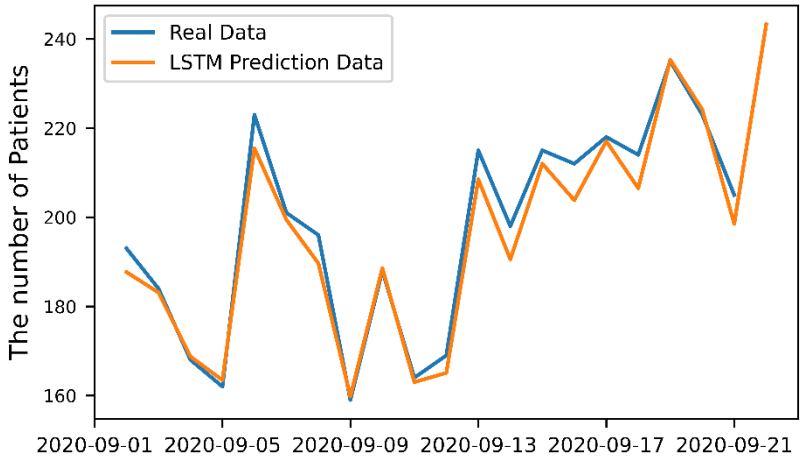

Fig10b Hybrid LSTM results of $f(x_{d1,t-1}, x_{d3,t-1}, y_{t-1}, c)$.

The summary of the model results can be seen in Table 3. As seen from the table adding air pollution data for the prediction increased the R^2 score in both ANN and LSTM between [0.01- 0.10]. When we used the wavelet transformed version of the air pollution data the R^2 values are also increased by max 0.28. When we remove $x_{d2,t-1}$ from the inputs of M.3 to create M.4, the LSTM train and test results are increased but the ANN train result is decreased.

Table 3. Model Evaluation Results

Model	Function of the Model	R^2 Train	R^2 Test	sMAPE % Train	sMAPE % Test
M.1 ANN	$f(y_{t-1}, y_{t-2}, c)$	0.770	0.108	27.506	17.823
M.1 LSTM*	$f(y_{t-1}, y_{t-2}, c)$	0.855	0.654	27.384	11.909
M.2 ANN	$f(x_{t-1}, y_{t-1}, y_{t-2}, c)$	0.879	0.117	26.410	17.045
M.2 LSTM	$f(x_{t-1}, y_{t-1}, y_{t-2}, c)$	0.881	0.636	26.153	12.153
M.3 Hybrid ANN	$f(x_{d1,t-1}, x_{d2,t-1}, x_{d3,t-1}, y_{t-1}, c)$	0.904	0.166	25.543	16.192
M.3 Hybrid LSTM*	$f(x_{d1,t-1}, x_{d2,t-1}, x_{d3,t-1}, y_{t-1}, c)$	**0.830**	**0.913**	26.043	**5.910**
M.4 Hybrid ANN	$f(x_{d1,t-1}, x_{d3,t-1}, y_{t-1}, c)$	0.891	0.202	24.166	15.641
M.4 Hybrid LSTM*	$f(x_{d1,t-1}, x_{d3,t-1}, y_{t-1}, c)$	**0.860**	**0.958**	26.618	**3.641**

Each different model uses a different function to find the y_t output. The models and accuracy metrics are calculated using the normalized data of variables. To find the predicted real data we should denormalize the y_t outputs. The prediction of the models can be seen in Table.4.

Table 4. Prediction values

Date (d.m.y)	Real Case Numb.(#)	Predicted COVID Case Numbers (#)							
		M.1 ANN	M.1 LSTM	M.2 ANN	M.2 LSTM	M.3 ANN	M.3 LSTM	M.4 ANN	M.4 LSTM
19.09.2020	218	231	218	221	224	201	202	215	**206**
20.09.2020	214	225	230	223	219	230	240	231	**235**
21.09.2020	235	204	212	206	204	210	226	213	**224**
22.09.2020	223	211	228	218	214	202	229	202	**227**
23.09.2020	205	222	236	222	219	264	237	257	**235**

The number of patients is related to all details. But there is a not significant relation. Pollution concentration is related with all details with alpha, "0.10". The role of mesoscale factors is less than other factors for both variables. Energy is transferred from level d3 (small scales) to d2 (meso scales) after pollution data. For hybrid ANN, to increase the performance of patient estimation, as weighting factors, small and large scale coefficients (d3 and d1) were considered as inputs in addition to the number of pollution concentrations.

CONCLUSION

Our results suggest the potential for substantial benefits from reducing air pollution exposure, even at relatively low fine particulate air pollution levels. Using air pollution data to predict the corona case number in İstanbul for 3 month period, increased the R2 score between [0-0.10]%. Nevertheless, the actual number of patients is influenced by many additional factors such as the country's health system. After error analysis, sMAPE [3.6-5.9] changed, the evidence of model results (M3, M4 Hybrid LSTM) and observation is sufficient. ($0.91<R^2<0.96$, alpha=<0.01). Performance of hybrid model 10% better than simple ANN model.

A lesson from our environmental perspective of the COVID-19 pandemic is that the quest for effective policies to reduce anthropogenic emissions, which cause both air pollution and climate change, needs to be accelerated. The COVID-19 pandemic ends with the vaccination of the population or with herd immunity through extensive infection. However, there are no vaccines against poor air quality and climate change. The remedy is to mitigate emissions. The transition to a green economy with clean, renewable energy sources will further both environmental and public health locally through improved air quality and globally by limiting climate change.

ACKNOWLEDGEMENT

Authors would like to thank for their data support to Ministry of Environment and Urbanization and Istanbul Metropolitan Municipality (IMM). The data can be taken from Turkish Health Ministry web page and Ministry of Environment and Urbanization and Istanbul Metropolitan Municipality.

REFERENCES

[1] Pozzer, A., Dominici, F., Haines, A., Witt, C., Münzel, T., & Lelieveld, J. (2020). Regional and global contributions of air pollution to risk of death from COVID-19. Cardiovascular research, 116(14), 2247-2253.

[2] Cohen AJ, et al. Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. Lancet 2017;389:1907–1918.

[3] Burnett R, et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. Proc Natl Acad Sci USA 2018;115: 9592–9597.

[4] Lelieveld J, et al. Comparison of mortality from ambient air pollution with other risk factors: a worldwide perspective. Cardiov Res 2020;doi: 10.1093/cvr/cvaa025.

[5] Cui Y, Zhang Z-F, Froines J, Zhao J, Wang H, Yu S-Z, Detels R. Air pollution and case fatality of SARS in the People's Republic of China: an ecologic study. Environ Health 2003;2:15.
[6] Driggin E,et al. Cardiovascular considerations for patients, health care workers, and health systems during the COVID-19 pandemic. J Am Coll Cardiol 2020;75: 2352–2371.
[7] Miller MR. Oxidative stress and the cardiovascular effects of air pollution. Free Radic Biol Med 2020;151:69–87.
[8] Wang B, et al. An effect assessment of airborne particulate matter pollution on COVID-19: a multi-city study in China. MedRxiv 2020;https://doi. org/10.1101/2020.04.09.20060137.
[9] Yao Y, Pan J., Wang W, Liu X, Kan H., Meng X, Wang W. Spatial correlation of particulate matter pollution and death rate of COVID-19. MedRxiv 2020;https://doi.org/ 10.1101/2020.04.07.20052142.
[10] Conticini E, Frediani B, Caro D. Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? Environ Poll 2020; 261:114465.
[11] Wu X, Nethery RC, Sabath MB, Braun D, Dominici F. Exposure to air pollution and COVID-19 mortality in the United States. MedRxiv 2020; doi: 10.1101/2020.04. 05.20054502
[12] Tadano, Y. S., et al. (2021). Dynamic model to predict the association between air quality, COVID-19 cases, and level of lockdown. Environmental Pollution, 268, 115920.
[13] Magazzino, C., Mele, M., & Schneider, N. (2020). The relationship between air pollution and COVID-19-related deaths: An application to three French cities. Applied Energy, 279, 115835.
[14] Jiang, Y., & Xu, J. (2021). The association between COVID-19 deaths and short-term ambient air pollution/meteorological condition exposure: a retrospective study from Wuhan, China. Air Quality, Atmosphere & Health, 14(1), 1-5.
[15] Shaffiee Haghshenas, S., et al. (2020). Prioritizing and analyzing the role of climate and urban parameters in the confirmed cases of COVID-19 based on artificial intelligence applications. International journal of environmental research and public health, 17(10), 3730.
[16] Bashir, M. F., et al. (2020). Environmental pollution and COVID-19 outbreak: insights from Germany. Air Quality, Atmosphere & Health, 13(11), 1385-1394.
[17] Velásquez, R. M. A., & Lara, J. V. M. (2020). Gaussian approach for probability and correlation between the number of COVID-19 cases and the air pollution in Lima. Urban Climate, 33, 100664.
[18] Huang, G., & Brown, P. E. (2021). Population-weighted exposure to air pollution and COVID-19 incidence in Germany. Spatial statistics, 41, 100480.
[19] Anderson, J. O., Thundiyil, J. G., & Stolbach, A. (2012). Clearing the air: a review of the effects of particulate matter air pollution on human health. Journal of medical toxicology, 8(2), 166-175.
[20] Santibañez, D. A., Ibarra, S., Matus, P., & Seguel, R. (2013). A five-year study of particulate matter (PM2. 5) and cerebrovascular diseases. Environmental Pollution, 181, 1-6.
[21] Anjum, M. S., Ali, S. M., Subhani, M. A., Anwar, M. N., Nizami, A. S., Ashraf, U., & Khokhar, M. F. (2020). An emerged challenge of air pollution and ever-increasing particulate matter in Pakistan; a critical review. Journal of Hazardous Materials, 123943.
[22] Rastgeldi Doğan, T., Yeşilnacar, M. İ., & Cullu, M. A. (2018). Seasonal investigation of atmospheric desert dust affecting sanliurfa using modis satellite and hysplit model data. Sigma: Journal of Engineering & Natural Sciences/Mühendislik ve Fen Bilimleri Dergisi, 36(3).
[23] Dogan, T. R., & Yalcin, S. P. (2020). The Atmospheric Transported Desert Dust Over Sanliurfa (Turkey) and its Structural Properties. Sigma: J. Eng. Nat. Sci, 38, 1837-1848.
[24] Awan, A. U., Sharif, A., Hussain, T., & Ozair, M. (2017). Smoking model with cravings to smoke. Advanced Studies in Biology, 9(1), 31-41.
[25] Hussain, T., Awan, A. U., Abro, K. A., Ozair, M., & Manzoor, M. (2021). A mathematical and parametric study of epidemiological smoking model: a deterministic stability and optimality for solutions. The European Physical Journal Plus, 136(1), 1-23.
[26] On the optimal control of coronavirus (2019-nCov) mathematical model; a numerical approach. Advances In Difference Equations Volume: 2020 Issue: 1 Article Number: 528 Published: SEP 25 2020
[27] Mathematical modeling of COVID-19 epidemic with effect of awareness programs. Infectious Disease Modelling 6 (2021) 448-460
[28] Muhammad Ozair, Takasar Hussain, Mureed Hussain, Aziz Ullah Awan, Dumitru Baleanu, Kashif Ali Abro, A Mathematical and Statistical Estimation of Potential Transmission and Severity of COVID-19: A Combined Study of Romania and Pakistan, BioMed Research International, Volume 2020, Article ID 5607236, 14 pages, https://doi.org/10.1155/2020/5607236
[29] Ozair, M., Hussain, T., Hussain, M., Awan, A. U., Baleanu, D., & Abro, K. A. (2020). A mathematical and

statistical estimation of potential transmission and severity of COVID-19: A combined study of Romania and Pakistan. BioMed research international, 2020.
[30] Analysis of fractional COVID-19 epidemic model under Caputo operator. Math Meth Appl Sci. 2021. https://doi.org/10.1002/mma.7294
[31] A fuzzy-based strategy to suppress the novel coronavirus (2019-ncov) massive outbreak. Appl. Comput. Math., v.20, n.1, special issue, 2021, pp.160-176
[32] Aziz Ullah Awan, Attia Sharif, Kashif Ali Abro, Muhammad Ozair, Takasar Hussain, Dynamical Aspects of Smoking Model with Cravings to Smoke, Nonlinear Engineering, (2021), https://doi.org/10.1515/nleng-2021-0008
[33] Takasar Hussain, Aziz Ullah Awan, Kashif Ali Abro, Muhammad Ozair, Mehwish Manzoor, A mathematical and parametric study of epidemiological smoking model: a deterministic stability and optimality for solutions, European Physical Journal Plus, 136:11, (2021), https://doi.org/10.1140/epjp/s13360-020-00979-4
[34] Şahin, M. (2020). The Association Between Air Quality Parameters and COVID-19 in Turkey. Pharmaceutical and Biomedical Research, 6, 49-58.
[35] Ali, H., Yilmaz, G., Fareed, Z., Shahzad, F., & Ahmad, M. (2021). Impact of novel coronavirus (COVID-19) on daily routines and air environment: evidence from Turkey. Air Quality, Atmosphere & Health, 14(3), 381-387.
[36] Turkish Health Ministry; https://covid19.saglik.gov.tr/TR-68443/covid-19-durum-raporu.html
[37] İTÜ Meteorology Department; http://kmg.itu.edu.tr/en/homepage
[38] Merry, R. J. E. (2005). Wavelet theory and applications: a literature study. DCT rapporten, 2005.
[39] Rhif, M., Ben Abbes, A., Farah, I. R., Martínez, B., & Sang, Y. (2019). Wavelet transform application for/in non-stationary time-series analysis: a review. Applied Sciences, 9(7), 1345.
[40] Chang, G. W., Lin, Y. L., Liu, Y. J., Sun, G. H., & Yu, J. T. (2021). A Hybrid Approach for Time-Varying Harmonic and Interharmonic Detection Using Synchrosqueezing Wavelet Transform. Applied Sciences, 11(2), 752.
[41] Zhang, L., Li, Z., Kirikkaleli, D., Adebayo, T. S., Adeshola, I., & Akinsola, G. D. (2021). Modeling CO 2 emissions in Malaysia: an application of Maki cointegration and wavelet coherence tests. Environmental Science and Pollution Research, 1-15.
[42] Kleene, S. C. (1956). Realization of nerve nets and finite automata. In Automata Studies (pp. 3-41). Princeton Univ. Press.
[43] Le, Q. V. (2013, May). Building high-level features using large scale unsupervised learning. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 8595-8598). IEEE.
[44] Lucchese, L. V., de Oliveira, G. G., & Pedrollo, O. C. (2021). Investigation of the influence of nonoccurrence sampling on landslide susceptibility assessment using Artificial Neural Networks. Catena, 198, 105067.
[45] Rostami, S., Toghraie, D., Shabani, B., Sina, N., & Barnoon, P. (2021). Measurement of the thermal conductivity of MWCNT-CuO/water hybrid nanofluid using artificial neural networks (ANNs). Journal of Thermal Analysis and Calorimetry, 143(2), 1097-1105.
[46] Hochreiter, S., & Schmidhuber, J. (1997). LSTM can solve hard long time lag problems. Advances in neural information processing systems, 473-479.
[47] Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: Continual prediction with LSTM.
[48] Voelker, A. R., Kaji, I., & Eliasmith, C. (2019). Legendre memory units: Continuous-time representation in recurrent neural networks.
[49] Devore, Jay L. (2011). Probability and Statistics for Engineering and the Sciences (8th ed.). Boston, MA: Cengage Learning. pp. 508–510
[50] Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. International journal of forecasting, 9(4), 527-529.
[51] Gelman, A., Goodrich, B., Gabry, J., & Vehtari, A. (2019). R-squared for Bayesian regression models. The American Statistician.