Prediction of Air Quality Index by Extreme Learning Machines

Burhan BARAN
Directorate of Environment and Urbanization
Malatya, Turkey
burhanbaran@gmail.com

***Abstract*—In this study, it is aimed to predict the Air Quality Index (AQI) by the Extreme Learning Machines (ELM) algorithm. For this purpose, six parameters have been selected which can affect the AQI. These are temperature, humidity, pressure, wind speed, PM_{10} and SO_2 respectively. First of all, "Forecast Sheet" application that presented in the Excel environment and the correlation analysis were used to determine the relationship status between these six parameters and AQI. Thus, the linearequations between each parameter and AQI were determined. Also, R values were determined for relationrate. Then, a training data set was created. 120 values belong to March, June, September and December of 2017, which includes six parameters, was determined as training data. AQI values of these six parameters were obtained by "AQI Calculator" application. The obtained AQI values were classified mathematically from 1 to 6 considering the AQI limits determined by National Air Quality Monitoring Network of Ministry of Environment and Urbanization. The first 12 day parameters of January, March, September and December of 2018 year are determined as test data. Then, ELM used 108 of the 120 generated data as training data and the remaining 12 of the data as test data. 10-fold cross validation method was used for the reliability of the result. Sine, sigmoid and hardlimit activation codes were used for classification. 6 different neurons were studied. As a result of cross-validation, the activation code and the number of neurons that achieved the highest success were applied to the actual test data. Finally, the classification values predicted by ELM and the mathematical classification values were compared. Thus, the success of the ELM was measured.**

***Index Terms*—Air Quality Index, SO_2, PM_{10}, Correlation, Extreme Learning Machine.**

I. INTRODUCTİON

Everyone needs the environment to survive. The environment is the place for living. In other words, the environment is a biosphere-living sphere where conditions are suitable for life. In the developing world, the increase of industry and human population poses a risk for environmental pollution [1]. Pollutants that can be present in the atmosphere as dust, smoke, gas and impure water vapor are called air pollution, which will have a negative impact on the health of people and living things. The problem of air pollution is closely related to the geographical, demographic and socio-economic characteristics of the regions. Therefore, one of the primary source areas of air pollution are cities due to the use of fossil fuels. Considering that the cities have both a large population and the population increases with migration, more and more individuals are expected to be affected by air pollution [2].

Particulate matter (PM) and sulfur dioxide (SO_2) are important pollutants. In addition, meteorological factors such as wind speed, direction, temperature, pressure, humidity are important for the distribution of air pollutants. PM is a very small particulate solid or liquid particles that are emitted in the atmosphere, except for large particles that can rapidly collapse due to their weight in the atmosphere. SO_2 is colorless, stifling and a acidic gas. SO_2 is released by the combustion of sulfur compounds found in the natural structure of coal and fuel oil [3]. PM_{10}, which is the most important part of the atmospheric particulate matter, is defined as particulate matter having a aerodynamic diameter of 10 μm or less. Particles with a size of 10 microns or less (PM_{10}) cannot be attached to the mucosa and the siller by the passage of harmful substances in the body, thus settling into the bronchi and the alveoli. Therefore, PM_{10} is considered to be the fraction that represents the particulate matter entering the human body through respiration and where the potential harmful elements are more connected [4]. Many countries are working to reduce air pollution through both national and international agreements. Turkey is one of the countries trying to take theseprecautions. For this purpose, Air Quality Measurement Stations have been established in many cities throughout the country, where the Ministry of Environment and Urbanization is authorized. Hourly data from these stations are collected and analyzed [5]. The Regulation on Air Quality Assessment and Management, published in the Official Newspaper No 26898 dated June 6, 2008, has been enacted in order to collect adequate information on air quality and to inform the public through warning thresholds [4].

In this study, the value of AQI is predicted by ELM taking into account the parameters of temperature, humidity, pressure, wind speed, PM_{10} and SO_2. In order to determine the effect of these six parameters on the value of AQI, "Forecast Sheet" application was used in Excel. Correlation analysis was used to obtain the linear equation of the current and predicted data obtained. Thus, prediction equations and R values of these equations were obtained. Then, using the "AQI Calculator" [6] application, the values of the AQI corresponding to these prediction values were determined. The accuracy of the results obtained by "AQI Calculator" application has been tested and verified with the current date parameters. The values obtained from these are confirmed by the available data on the implementation of the National Air Quality Monitoring Network of the Ministry of Environment and Urbanization. Then the obtained values of the AQI were classified as 1 to 6 mathematically considering the AQI classification determined by the National Air Quality Monitoring Network of the Ministry of Environment and Urbanization.

120 of the 122 data of March, June, September and December of 2017 were used as training data. 108 of these 120 data were used as training and the remaining 12 were used as test data. 10-fold cross validation method was used to make the results reliable. The activation code and the number of neurons, which give the correct results at the highest rate and in the shortest time, have been applied to the actual test data.The data to be tested are the first 12 data of January, March, September and December of 2018. It is aimed by ELM to correctly predict the class of these four different 12-day data. Finally, the predicted and mathematical classification results were compared.

II. SCIENTIFIC STUDY REVIEW

PM_{10} concentrations in provinces of Turkey's Eastern and Southeastern Anatolia Region were evaluated by A. Dolar and friends in 2014. The data are taken from the Ministry of Environment and Urbanization National Air Monitoring Stations. PM_{10} data were analyzed spatially and temporally in the light of statistical analysis. Compared to national and international limits, the daily average PM_{10} values obtained from 23 stations in 2014 were compared. PM_{10} values of each station were evaluated statistically. During the temporal evaluation of the concentrations, two groups were examined as warm-up and non-warm period. As a result of the study, the average of the PM_{10} concentrations for all provinces, the average of warming period and the average of non-warming period were calculated as 56, 66 and 47 $\mu g\ m^{-3}$, respectively. All data were found to exceed 40 $\mu g\ m^{-3}$ that is 2019 target of Turkey [4].

Gocheva-Ilieva et al. conducted a study on the Box–Jenkins method and factor analysis to examine the primary concentrations in the Blagoevgrad region of Bulgaria, which were measured hourly during a 1-year period and caused air pollution. Using factor analysis, multiple linearity was determined between six pollutants. Pollutants are grouped in 3 factors. The contribution of these factors to total pollution was determined [7].

Souza et al. conducted a study to determine the relationship between air pollutant concentrations and respiratory causes in children. Ecological time series method was used. The relationship between hospital admissions and daily air pollutant concentrations of children under 6 years of age was analyzed. Poisson regression and generalized model analysis techniques are used for statistical analysis. These analysis techniques are complemented by better quality predictions. For each 10.49 $g\ /\ m^3$ increase in pollutant PM_{10} levels, the relative risk generalized additive model in the general seasonal additive model of the main components was increased by 3% while the increase rate was 2% [8].

A study was conducstedon the prediction of the concentration of air pollutants by Zhang et al. For this purpose, feed forward neural network propagation (FFANN-BP) and Extreme Learning Machines methods are compared. Compared to FFANN-BP, which is based on gradient learning algorithm, ELM has achieved many important achievements. ELM has shown good performance in precision, robustness and generalization. There was no significant difference between the prediction accuracies of both models. ELM has been shown that it has the best performance in predictive indicators such as R-square and RMSE [9].

Zhu et al., discussed air quality predicts using machine learning approaches to predict air pollution (ozone, particulate matter ($PM_{2.5}$) and SO_2) hourly concentrations. For this purpose, machine learning method has been developed. Refined models are proposed for prediction of hourly air pollution concentration. These predictes were considered as multi-tasklearning (MTL) problem and the previous 24-hour meteorological data were used. The problem is formulated as normalized MTL. Advanced optimization methods have been used to solve different formulas. As a result, the proposed method yields much better results [10].

A study has been conducted on the effect of meteorological parameters on air pollution in Keçiören district of Ankara province by Oğuz et al. The parameters used are temperature, relative humidity, pressure, wind speed, SO_2 and PM_{10}. The correlation coefficients between the values of these parameters were calculated to calculate the relationship between the parameters. 2013 and 2014 data were used in the study. At the end of the study, changes in PM_{10} and SO_2 values were examined. It has been determined that air pollution increases during the seasons when fuels are used. According to the results obtained, PM_{10} was seen most in winter and affected by pressure and wind speed. SO_2 was also seen in winter and was only related to relative humidity [3].

In the study conducted by Cetinyokus, it is aimed to determine the possible effects of air pollutants in Dilovası district of Kocaeli province by ALOHA software. The data of 2010 and 2017 were used in the study. The data were obtained from the National Air Quality Monitoring Network, which was established by the Ministry of Environment and Urbanization. Modeling studies were conducted directly from the source and instant ejection scenario. In 2017, it was observed that concentrations for CO and O_3 air pollutants decreased. Although all air pollutants are in a good class of air stability class, it has been determined that they may cause serious effects on the instant ejection scenario. All air pollutants have been determined that cause toxic effects. The prevailing wind direction in the region has been found advantageous in the direction of directing toxic hazard zones to less populated areas [11].

Dursun et al. investigated the relationship between sulfur dioxide (SO_2) concentrations and airborne pollutants in the air at Baku city center. Data from 2012 and 2013 were used. First,

SO_2 analysis was performed. Then the relations between the concentration values and meteorological data were examined. The relationship between wind, precipitation, humidity, temperature, and SO_2 was investigated. As a result of review, the highest SO_2 value was 0,035 mg/m^3 daily. According to the correlation tests with each parameter, the relations were negative [12].

In this study, temperature, humidity, pressure, wind speed, PM_{10} and SO_2 parameters were used. According to these data, the predicted classification results of the AQI predicted by ELM and the results of the mathematical AQI classification were compared. Thus, the success of the ELM was measured. This study on the classification of AQI is thought to contribute to the literature.

III. CORRELATION ANALYSIS

Correlation analysis is one of the most frequently used statistical methods to determine the association and direction between two variables. The correlation coefficient (R) is a ratio, ranging from -1 to +1. If this coefficient is positive, one of the variables increases and the other increases; If negative, one of the variables increases and the other decreases. The Pearson correlation coefficient is calculated by equation 1 [13].

$$R = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}} \quad (1)$$

x and y show the characteristics of two continuous structures measured from the same individuals [14].

In order to determine the ratio of temperature, humidity, pressure, wind speed, PM_{10} and SO_2 parameters to AQI, the Excel "Forecast Sheet" application has been used. Correlation analysis was performed through this application. Thus, linear equations between each parameter and AQI were obtained. Also, according to the obtained equation, R values were obtained to determine the relationship rate between the parameters.

A. Correlation Analysis Between AQI andParameters

The distribution graphs of 120 temperature and AQI values for January, March, September and December of 2017 are as in figure 1 (a). The R value indicating the relationship between these two parameters was obtained as -0,359. Accordingly, it is concluded that there is a negative relationship between these two parameters. The distribution graph corresponding to one hundred twenty pieces of humidity and AQI values is as in figure 1 (b). The R value indicating the relationship between humidity and AQI was 0,558. There is a positive directional relationship. The distribution graph for 120 values between January, March, September and December of 2017 between the pressure and AQI was obtained as in figure 1 (c). The R value indicating the relationship between these parameters was obtained as 0,511. There is a positive directional relationship. The dispersion graph of 120 data corresponding to wind speed and AQI values was obtained as in figure 1 (d). The R value indicating the relationship between these two parameters was obtained as -0,441. So there is a negative relationship. The PM_{10} values are taken from the National Air Quality Monitoring Network of the Ministry of Environment and Urbanization. PM_{10} values are 24 hours average values. The distribution graphs of 120 PM_{10} and AQI values for January, March, September and December of 2017 are as in figure 1 (e). The R value showing the relationship between PM_{10} and AQI was 0,996. There is a positive directional relationship. SO_2 values were taken from the National Air Quality Monitoring Network of the Ministry of Environment and Urbanization.SO_2 values are 24 hours average values. The distribution graph corresponding to one hundred twenty SO_2 and AQI values is as in figure 1 (f). The R value indicating the relation between these parameters was 0,394. There is a positive directional relationship.

According to this, the highest correlated parameter with AQI was PM_{10} with R=0,990. Then, the highest relationship rates were respectively humidity with R=0.558, pressure with R=0.511, wind speed with R=-0.441, SO_2 with R=0.394 and temperature R=-0.359. In addition, all linear equations are as figure 1.

IV. RECOMMENDED METHOD

A. Extreme Learning Machines

This section firstly describes the extreme learning machine developed by Huang et al [15]. Then, a study was conducted on how to obtain the predicted classification of the AQI value corresponding to the temperature, humidity, pressure, wind speed, PM_{10} and SO_2 data which are selected in the algorithm. Machine Learning makes inferences from ready data by using statistical and mathematical methods. And it is a method of predicting the unknown value through these inferences. ELM is a single-layer and feed-forward artificial neural network model. The input weights in ELM are randomly selected; however, the output weights are calculated analytically. Therefore, ELM shows faster learning speed and better generalization performance than traditional learning algorithms. While the linear function is used in the output layer, the hidden layer uses activation functions such as sigmoidal, sine and hard-limit [16].

Parameterssuch as thresholds and weights in traditional feed-forward neural Networks must be updated through gradient-based learning algorithms. Nevertheless, the learning process required to achieve a good result may take time. The fault can be attached to a local point. In addition, input-output weights and threshold values do not affect the performance of the network in a single hidden layer feed-in network. In figure 2, a single hidden layer and a feed-forward artificial neural network is shown [17]. The target of the neuron network is to find the function *f*: $X \rightarrow Y$. A single-layer feed forward neural network with *N* hidden node is defined as in equation 2 [18].

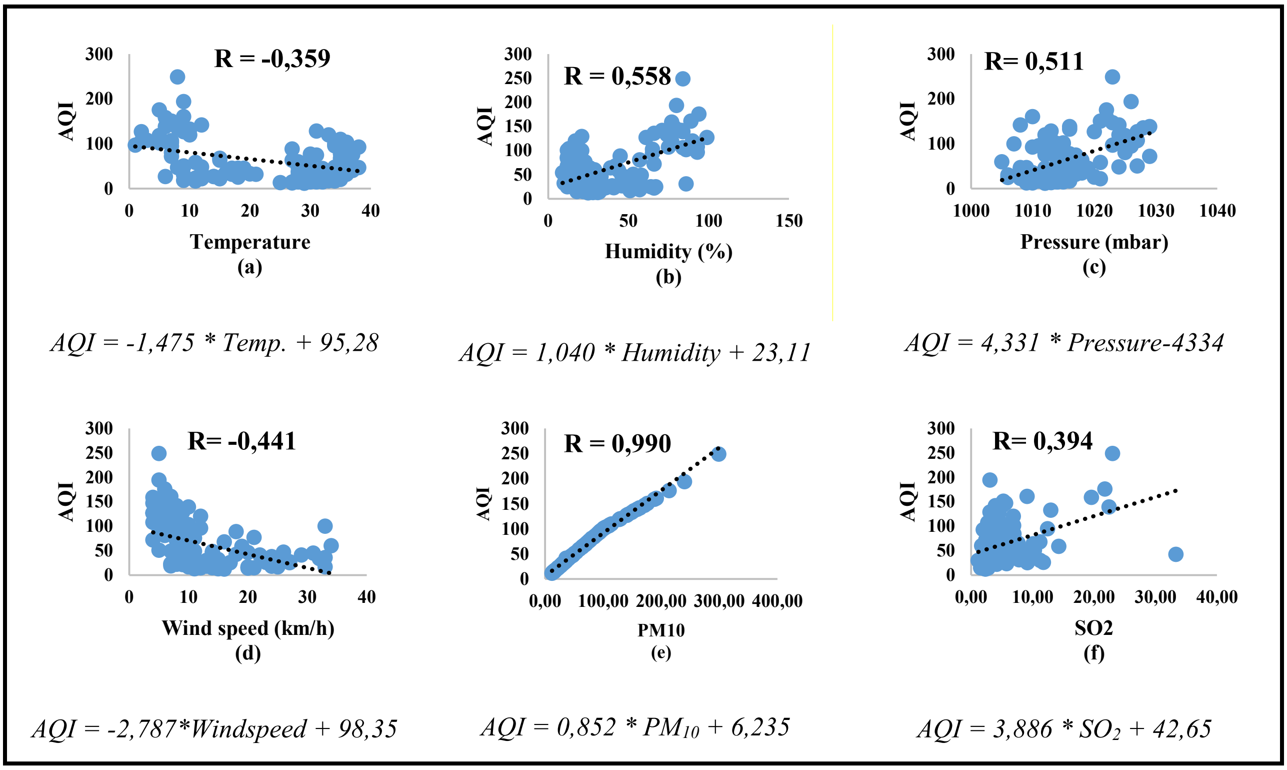

Fig. 1. Correlation Analysis

a) AQI-Temperature b) AQI-Humidityc) AQI-Pressure d) AQI-Wind speed e) AQI-PM_{10} f) AQI-SO_2

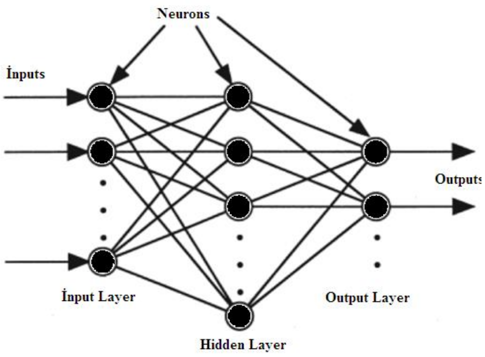

Fig. 2. Feed-forward Artificial Neural Network [17].

$$f_N(x) = \sum_{i=1}^{N} B_i G(a_i, b_i x), x \in R, a_i \in R \quad (2)$$

where a_i and b_i are the learning parameter, B_i is i. weight of the hidden node. For a feed-forward network with a generalized single layer, the output function of the ELM is as in equation 3.

$$f_N(x) = \sum_{i=1}^{N} B_i G(a_i, b_i x) = \beta \times h(x)) \quad (3)$$

B. Training Data Set forELM

The PM_{10} and SO_2 parameters were used to construct the training data set. In addition to the values of these parameters, NO_X value on “AQI Calculator” application has been selected as 1 and AQI values have been obtained. For the confirmation that the AQI values calculated by the “AQI Calculator” application were calculated correctly, it was compared with the instantaneous AQI values corresponding to the immediate PM_{10} and SO_2 values received from the Ministry of Environment and Urbanization's National Weather Monitoring Network website [19]. As a result of comparison, very close values were observed. Thus, it has been confirmed that the “AQI Calculator” application performs the correct calculation.

Then, the data of temperature, humidity, wind speed and pressure for the months of January, March, September and December were taken from the website of www.timeanddate.com [20]. The data that taken from this site belong to 12:00 hours. It was taken 29 days from March, 30 from June, 30 from September and 31 days from December. The reason for the reduction of 2 days from March is to obtain 120-day data for 10-fold cross validation. Thus, training data consisting of 1 AQI data were obtained in response to 6 input data consisting of 120 lines of temperature, humidity, pressure, wind speed, PM_{10} and SO_2. The data obtained from 120 AQI were classified as 1 to 6 mathematically. As a result of this process, 1st class consisted of 66, 2nd class consisted of 26, 3rd class consisted of 22, 4th class consisted of 5, 5th class consisted of 1 data. There is no AQI value with the 300-500 range within the present AQI data. Therefore, 6th class did not occur.

V. CASE STUDY

Three different activation functions, sine, sigmoidal and hardlimit were used by ELM for the classification of training and test data. Test times and test accuracy rates of these activation codes were compared. These values are given in Table I. The number of hidden neurons used in this initial comparison phase is 108. As the training data consisted of 108 lines, a maximum of 108 neurons could be used. A series of computer simulations were carried out on MATLAB environment. In ELM, when the output weights are calculated analytically, the input weights are calculated randomly. Therefore, the results of each program run close to each other but different values. To overcome this problem 10-fold cross validation technique is used in evaluation of the ELM. The achievements were measured by “Test Time” and “Test Accuracy” values.

TABLE I. TEST TIMES VE TEST ACCURACY VALUES AT DIFFERENTACTIVATION FUNCTIONS (100 NEURON)

Activation Function	Test Time (s)	Test Accuracy (%)
Sig	0,0017	58,33
Sin	0,0017	16,67
Hardlim	0,00072	66,67

When the test accuracy rates are examined according to the ELM activation functions, it is seen that the highest value is met by the hardlimit function with 66,67% test accuracy rate and 0,00072 second test period. However, in order to investigate how it reacts to its different functions in different neuron counts, the dependence of these 3 activation functions on 6 different neurons are investigated. After the study, the values in Table II were obtained.

As can be seen from Table II, the highest test accuracy rate was 74,17% (75%) and hardlim activation function and 50 neuron counts. The highest test accuracy rates were in hardlim-108 and sig-70 neuron numbers and they were 71,67% and 70,83%, respectively. According to the results obtained in the study for the test data to be composed of 12 lines, the hardlim-50 activation function was used. Test data consisting of 12 lines of 4 different periods were studied. Training data were monthly data for March, June, September and December of 2017, while test data were collected from January, March, September and December of 2018 for the first 12 days. PM_{10} and SO_2 values are taken from the National Weather Monitoring Network of the Ministry of Environment and Urbanization. Temperature, humidity, pressure and wind speed data are taken from the website www.timeanddate.com.

TABLE II. TEST TIMES VE TEST ACCURACY VALUES AT DIFFERENT ACTIVATION FUNCTIONS AND DİFFERENT NEURON NUMBERS

Activation code and hidden neuron	Test Time (s)	Average Test Accuracy (%)
Sig-5	0,0004340	52,50
Sin-5	0,0003412	22,50
Hardlim-5	0,0004032	55,00
Sig-20	0,0003233	52,50
Sin-20	0,0003271	25,83
Hardlim-20	0,0003893	60,00
Sig-50	0,0003748	60,00
Sin-50	0,0003624	33,33
Hardlim-50	***0,0004156***	***74,17***
Sig-70	**0,0003687**	**70,83**
Sin-70	0,0003682	25,00
Hardlim-70	0,0004052	60,00
Sig-90	0,0003932	67,50
Sin-90	0,0003884	30,83
Hardlim-90	0,0004133	68,33
Sig-108	0,0004263	73,33
Sin-108	0,0004406	19,17
Hardlim-108	**0,00004607**	**71,67**

The limit values of the AQI taken from the Ministry of Environment and Urbanization's website of the National Weather Monitoring Network are as in Table III.

TABLE III. AQI LIMIT VALUES AND ATTENDED CLASS [19]

AQI VALUES					
0-50	50-100	100-150	150-200	200-300	300-500
1	2	3	4	5	6

TABLE IV. PARAMETERS VALUES AND MATHEMATICALLY CALCULATED CLASSIFICATIONS FOR 12.01.2018-01.01.2018 DATE

Temperature	Humidity (%)	Pressure (mbar)	Wind speed (km/h)	PM_{10}	SO_2	Calculated AQI	Mathematical Class
9	50	1020	7	105,72	20,01	103	3
7	61	1021	9	91,95	20,79	92	2
10	41	1025	9	32,56	14,29	33	1
12	41	1025	7	55,17	16,30	55	2
11	52	1029	6	96,59	15,97	97	2
11	54	1029	7	47,00	12,15	47	1
11	61	1020	6	63,24	10,86	63	*2*
8	75	1011	8	44,99	8,13	45	1
7	79	1012	11	64,78	15,35	65	*2*
7	57	1018	5	71,40	17,00	71	2
10	39	1014	20	28,62	11,82	29	1
10	61	1009	13	179,46	14,33	153	4

AQI= Air Quality Index

Accordingly, 6 input values for the ELM classification between 12.01.2018-01.01.2018 for the hardlim-50 activation code and the number of neurons, and the corresponding AQI and mathematical classification values for these values are obtained as in Table IV. The mathematical classification values corresponding to the computed AQI values vary between 1, 2, 3 and 4.

For the ELM classification between 12.03.2018 and 01.03.2018, 6 input values and the corresponding AQI and mathematical classification values are obtained as in Table V. The AQI values were 28 at the lowest and 140 at the highest. Mathematical classification values vary between 1, 2 and 3 values.

TABLE V. PARAMETERS VALUES AND MATHEMATICALLY CALCULATED CLASSIFICATIONS FOR 12.03.2018-01.03.2018 DATE

Temperature	**Humidity (%)**	**Pressure (mbar)**	**Wind speed (km/h)**	**PM_{10}**	**SO_2**	**Calculated AQI**	**Mathematical Class**
18	36	1015	9	115,14	10,56	110	3
17	34	1014	8	53,93	14,87	54	2
15	38	1015	14	31,59	10,86	32	1
14	51	1009	27	35,11	4,55	35	1
13	84	1007	11	159,84	6,84	140	3
12	74	1017	5	126,28	3,96	118	3
11	83	1013	9	122,63	10,41	115	3
17	41	1011	9	85,36	10,84	85	2
17	50	1008	27	75,83	5,77	76	2
16	55	1014	20	80,13	16,19	80	2
9	48	1019	7	46,90	23,82	47	1
7	79	1015	17	28,17	7,99	28	1

AQI= Air Quality Index

Temperature, humidity, pressure, wind speed, PM_{10}, SO_2 values measured in the 12.09.2018 - 01.09.2018 date and the AQI values calculated in response to these values are as in Table VI. In addition, the mathematical classification values corresponding to the calculated AQI values can also be seen in the table. The calculated values for this month are generally close to each other and range from 27 to 69.

TABLE VI. PARAMETERS VALUES AND MATHEMATICALLY CALCULATED CLASSIFICATIONS FOR 12.09.2018-01.09.2018 DATE

Temperature	**Humidity (%)**	**Pressure (mbar)**	**Wind speed (km/h)**	**PM_{10}**	**SO_2**	**Calculated AQI**	**Mathematical Class**
31	25	1019	13	65,42	4,94	65	2
28	31	1017	10	27,29	4,26	27	1
25	50	1015	12	27,26	1,67	27	1
29	24	1012	12	32,2	6,64	32	1
30	22	1012	10	34,66	9,89	35	1
29	30	1009	17	50,70	2,80	51	2
34	21	1005	25	69,21	29,05	69	2
35	13	1008	10	45,67	11,27	46	1
35	12	1009	10	43,69	11,52	44	1
34	22	1008	10	45,13	13,33	45	1
35	14	1011	11	43,06	15,02	43	1
34	17	1013	11	43,29	15,67	43	1

AQI= Air Quality Index

The values of temperature, humidity, pressure, wind speed, PM_{10}, SO_2 parameters between 12.12.2018-01.12.2018 are shown in Table VII. The AQI values calculated by the "AQI Calculator" application and the mathematical classification values are also included in this table considering PM_{10} and SO_2 values. It is seen that mathematical classification values ranged between 1 and 2.

TABLE VII. PARAMETERS VALUES AND MATHEMATICALLY CALCULATED CLASSIFICATIONS FOR 12.12.2018-01.12.2018 DATE

Temperature	**Humidity (%)**	**Pressure (mbar)**	**Wind speed (km/h)**	**PM_{10}**	**SO_2**	**Calculated AQI**	**Mathematical Class**
10	75	1010	19	46,07	8,16	46	1
14	57	1013	22	44,49	7,16	44	1
10	79	1017	6	91,73	15,10	92	2
10	68	1020	6	68,27	11,68	68	2
8	74	1017	4	55,61	16,06	56	2
10	84	1012	6	94,89	6,28	95	2
11	79	1011	9	97,22	22,12	92	2
11	85	1015	5	71,69	5,65	72	2
8	92	1019	7	84,67	8,44	85	2
11	69	1022	6	92,01	9,26	92	2
11	77	1024	6	67,91	6,42	68	2
13	80	1018	11	36,29	3,72	36	1

AQI= Air Quality Index

TABLE VIII. COMPARISION OF MATHEMATICALLY AND ELM CLASSIFICATION

12.01.2018-01.01.2018		**12.03.2018-01.03.2018**		**12.09.2018-01.09.2018**		**12.12.2018-01.12.2018**	
Mathematical Class	ELM Prediction Calass	Mathematical Class	ELM Prediction Calass	Mathematical Class	ELM Prediction Calass	Mathematical Class	ELM Prediction Calass
3	*2*	*3*	*2*	*2*	*1*	1	1
2	2	2	2	1	1	1	1
1	1	1	1	1	1	2	2
2	2	1	1	1	1	2	2
2	2	3	3	1	1	*2*	*1*
1	1	3	3	*2*	*1*	2	2
2	*1*	3	3	*2*	*1*	2	2
1	1	2	2	1	1	*2*	*1*
2	*1*	*2*	*1*	1	1	2	2
2	2	2	2	1	1	2	2
1	1	*1*	*2*	1	1	*2*	*1*
4	4	1	1	1	1	1	1

The mathematical classification results obtained from Table IV-V-VI-VII were compared with the predicted classification results obtained by the ELM algorithm. These results are shown in Table VIII.

Training data consisting of 108 lines of 6 parameters were applied to ELM algorithm. The values of the first twelve days of January, March, September and December of 2018 were predicted by ELM. For this purpose, 75% successful prediction was realized by ELM in 12.01.2018-01.01.2018 period. Three incorrect classification predicts were made. While the first one of them was between 100-150 values at 12.01.2018, ELM predicted the value of AQI at this date between 50-100 values. This error is considered to be due to the fact that at this date it is 103 which is very close to the limit value of AQI100 and ELM cannot predict it. At 06.01.2018-04.01.2018 the AQI value was between 50-100 values, while the ELM predicted 0-50 values.

During the period of 12.03.2018-01.03.2018, 9 of 12 data were correctly predicted by ELM. While the AQI value calculated on 12.03.2018 was between 100-150 values, incorrect prediction was made that according to the classification result made by ELM, the value of the AQI would be between 50-100values.It was calculated as 50-100 on 04.03.2018 and it was incorrectly predicted by ELM from 0-50. Likewise, while the AQI calculated on 01.03.2018 was between 0-50, the predicted AQI was between 50-100.

In the ELM classification for the prediction of the value of 12 of the values between 12.09.2018-01.09.2018, 75% performance was achieved. Three false predictions this month are on 12, 07 and 06 September. In all of the false predicts made in these three days, the value of AQI should be calculated between 50 and 100 and it was calculated as 0-50. It is thought that this is due to the fact that the calculated AQI values are close to 50 which is the limit value of all three values and that ELM cannot make an exact classification due to this reason. In the three false predicts made between 12.12.2018-01.12.2018, the AQI values could not be predicted between 50-100 values and were predicted to be 0-50values. As a result, 36 of the 48 data belonging to a total of 4 months were correctly predicted by ELM.

VI. RESULTS

In this study, it is aimed to predict the AQI values by the ELM algorithm. For this purpose, six parameters have been selected which can affect the AQI. These are temperature, humidity, pressure, wind speed, PM_{10} and SO_2 respectively. Firstly, correlation analysis was used to determine the relationship between these six parameters and the AQI, using the "Forecast Sheet" application in Excel. The linear equations between each parameter and AQI and the R values were obtained to determine the relationship rate. Then, a training data set was created. 120 value belong to March, June, September and December of 2017, which includes six parameters, was determined as training data. AQI values of these six parameters were obtained by "AQI Calculator" application. The values obtained from theseare confirmed by the available data on the implementation of the National Air Quality Monitoring Network of the Ministry of Environment and Urbanization.

The obtained values of the AQI were classified as 1 to 6 mathematically considering the AQI classification determined by the National Air Quality Monitoring Network of the Ministry of Environment and Urbanization. The first twelve day parameters of January, March, September and December of 2018 are determined as test data. Then, ELM was subjected to a total of 10 cross-validation procedures using 108 of the 120 data generated here as training data and the remaining 12 as test data. Sine, sigmoid and hardlimit activation codes were used in the classification. 6 different neurons were studied. As a result of this cross-validation, the highest performance was achieved with the activation code and the number of neurons applied to the actual test data.

When the test accuracy rates were examined according to the ELM activation functions, the highest value was met by the hardlimit function with 66,67% test accuracy rate and 0,00072 second test period. However, in order to examine how it responds to different neuron counts and other functions, the dependence of these 3 activation functions on 6 different neurons is investigated. After the study, the most suitable solution was found in the hardlim-50 function with 74,17% test accuracy and 0,0004156 test period. For the actual test data, data from the first 12 days of January, March, September and December of 2018 were used.Accordingly, for the hardlim-50 activation code and the number of neurons, the AQI mathematical classification values of 4 different seasons were obtained. Training data that consisting of 108 lines and belonging to 6 parameters were applied to ELM algorithm. The values of the first twelve days of January, March, September and December of 2018 were predicted with ELM. 36 of 48 data from 4 different periods were correctly predicted by ELM. Finally, the success of ELM was measured by comparing the AQI values predicted by ELM with the mathematical classification values.

REFERENCES

[1] Y. Kaplan, U. Saray, E. Azkeskin, "Hava Kirliliğine Neden olan PM_{10} ve SO_2 maddesinin Yapay Sinir Ağı kullanılarak Tahmininin Yapılması ve Hata Oranının Hesaplanması", Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 14:025201, pp.1-6, 2014.

[2] S. Menteşe, Ş. Tağıl, "Bilecik'te İklim Elemanlarının Hava Kirliliği Üzerine Etkisi", Balıkesir University The Journal of Social Sciences Institute, vol. 15:28, pp. 3-16, 2012.

[3] K. Oğuz, M. A. Pekin, "Meteorolojik Koşulların Hava Kirliliği Üzerine Etkilerinin İncelenmesi: Keçiören İlçesi Örneği", VII. Uluslararası Katılımlı Atmosfer Bilimleri Sempozyumu, 28-30 Nisan 2015.

[4] A. Dolar, H.T.K. Saraç, "Türkiye'nin Doğu İllerindeki Hava Kalitesinin PM_{10} Yönüyle İncelenmesi",Iğdır Üni. Fen Bilimleri Enst. Der., 5(4): 25-32, 2015.

[5] D. Turgut, İ. Temiz, "Time Series Analysis And Forecasting For Air Pollution In Ankara: A Box-Jenkıns Approach", alphanumericjournal, vol. 3, Issue 2, pp. 131-138, 2015.

[6] AQI Calculator, 2019, URL: https://app.cpcbccr.com/ccr_docs/AQI%20-Calculator.xls.

[7] S.G. Gocheva-Ilieva, A.V. Ivanov, D.S. Voynikova, D.T. Boyadzhiev, “Time series analysis and forecasting for air pollution in small urban area: an SARIMA and factor analysis approach”,Stochastic Environmental Research and Risk Assessment, vol. 28, Issue 4, pp. 1045–1060, 2014.
[8] J.B. Souza, V.A. Reisen, J.M. Santos, G.C. Franco, “Principal components and generalized linear modeling in the correlation between hospital admissions and air pollution”, Rev Saude Publica, vol. 48:3, pp. 451-458, 2014.
[9] J. Zhang, W. Ding, “Prediction of Air Pollutants Concentration Based on an Extreme Learning Machine: The Case of Hong Kong”, International Journal of Environmental Research and Public Health, Int. J. Environ. Res. PublicHealth, vol. 14:114, 2017. doi:10.3390/ijerph14020114.
[10] D. Zhu, C. Cai, T. Yangand, X. Zhou, “A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization”, Big Data Cogn. Comput.,vol. 2:5, 2018. doi:10.3390/bdcc2010005.
[11] S. Cetinyokus, “Determination of Possible Effects of Air Pollutants for the Kocaeli-Dilovasi”, Aksaray J. Sci. Eng. vol. 1, Issue2, pp. 121-133, 2017.
[12] Ş. Dursun, İ. İbrahimova, “Bakü Hava Kirlenmesinde SO_2’nin Rolü ve Meteorolojik Olaylarla İlişkisinin Araştırılması”, Avrupa Bilim ve Teknoloji dergisi, Cilt 1, No. 3, pp. 84-91, Haziran 2014.
[13] Tan, A. Korelesyon ve Regresyon Analizi, (Available at: www.ahmettan.com.tr/wp-content/uploads/2013/03/Bölüm9.ppt
[14] Taşdelen B. Korelasyon ve Regresyon Analizi, (Available at:https://docplayer.biz.tr/21378989-Istatistik-ii-korelasyon-ve-regresyon.html)
[15] G.B. Huang, Q.Y. Zhu, C.K. Siew, “Extreme Learning Machine: Theoryand Applications. Neurocomputing”,70:489-501, 2006.
[16] Y. Kaya, “A Fast Intelligent Diagnosis System For Thyroid Diseases Based On Extreme Learning Machine”, Bilim ve Teknoloji Dergisi A-Uygulamalı Bilimler ve Mühendislik, vol. 15:1, 2014.
[17] B. Baran, “Sınır Değerler Arasında Kalan Evsel Atıksu Numune Analizi Sonucunun Aşırı Öğrenme Makineleri İle Sınıflandırılması”, Journal of Engineering Sciences and Design, 7 (1), pp. 18-25, 2019.doi: 10.21923/jesd.457085.
[18] F. Ö. Çatak, “Rassal Bölümlenmiş Veri Üzerinde Aşırı Öğrenme Makinesi ve Topluluk Algoritmaları ile Sınıflandırma”, IEEE 23. Sinyal İşleme ve İletişim Uygulamaları Kurultayı(SIU), pp.16-19, 2015.
[19] National Air Quality Monitoring Network of Ministry of Environment and Urbanization, www.havaizleme.gov.tr, 2019.
[20] Malatya, Temperature, pressure, windspeed and humidity values, 2017. URL:https://www.timeanddate.com/weather/turkey/malatya/historic?month=1&year=2018, 2019.