CO2 Concentration Forecasting in an Office Using Artificial Neural Network

Mahsa Khorram, Pedro Faria, Omid Abrishambaf, Zita Vale, João Soares
Polytechnic of Porto (IPP), Porto, Portugal
makgh@isep.ipp.pt, pnf@isep.ipp.pt, ombaf@isep.ipp.pt, zav@isep.ipp.pt, jan@isep.ipp.pt

***Abstract*--Uncertainty is the state of all operation, components, and objective environment that makes impossible to describe the existing state. Forecasting techniques are essential in the field of knowledge development to overcome the uncertainty to increase the efficiency of all systems. In this paper, artificial neural network algorithm is applied to forecast the CO2 concentration in an office building. The algorithm is implemented in Rstudio software using neural net package. The case study of the paper presents two scenarios with different input data to propose the impacts of train data on forecasting algorithms results. The used dataset in the case study is real data that have been monitored for 2 years. The obtained results of algorithms show the predicted values of CO2 concentration in one office for 600 minutes of a working day. The mean percentage error means absolute percentage error, and standard deviation of predicted data for both scenarios are presented in results section.**

***Index Terms*— Artificial Neural Network, CO2, Forecasting.**

I. Introduction

Nowadays, increment of energy consumption and its environmental consequences have become as a concern for the world [1]. According to [2], 40% of total energy consumption and 36% of Carbon Dioxide (CO2) emissions are related to European buildings. Therefore, buildings are moving toward to be intelligent to have control on their conditions [1]. In power system area, the inherent uncertainty of environment can make a lot of problems [3]. Artificial Intelligence (AI) techniques are being used to overcome these uncertainty issues and provide solutions to deal with efficient program to manage the energy. AI also can be considered as an academic way to make computers and equipment capable of intelligent behavior [4].

One of the most appropriate AI techniques which is able to address several problems in power systems is Artificial Neural Networks (ANN). ANN is able to learn, train, and predict the data to deal with uncertainty. They use a method similar to the mechanism of human brain work through layers of neurons [5]. Since the ANN become one of the most important topics that the science concerns in to improve and adapt machines in every way to serve the human being. The applications of ANN increase every day because of its characteristics. ANN have been used in wide range applications such as process modeling and control, machine diagnostics, portfolio management, credit rating, targeted marketing, voice recognition, financial forecasting, intelligent searching, and medical diagnosis [6],[7].

Forecasting is important for planning and decision-making in all fields to predict the conditions of the problems before making any decision [8]. Therefore, many forecasting methods have been developed to produce accurate predicted values.

CO2 is the product of combustion, is considered as the result of human metabolism, and concentrations in the buildings should be determined for informing the amount of fresh air in the space. The high levels of CO2 concentration in close spaces can cause headaches and fatigue, and a higher concentration can cause more health issues such as nausea, dizziness, and vomiting. CO2 is a colorless, odorless and toxic gas, its invisibility, taste and smell impossibility make it a dangerous gas, because of its capability of killing and causing huge health damages. Carbon dioxide is considered to be an issue in the context of indoor air quality, because of its impact on health, and quality of air.

Some of health effects that could be a result of carbon dioxide damages have been mentioned below [9] - [11]:

- Low concentrations of this gas effect on the people with heart disease causing chest pain. The air environment in general could cause exhaustion and fatigue.
- Moderate concentrations of this gas could cause angina, reduced brain function and impaired vision.
- With the raising of gas concentrations, the Risk factors is increased causing more dangerous Syndrome such as: nausea, dizziness, confusion.

Forecasting methods have been employed in several researches for different purposes such as [12] which authors have been used ANN for real-time load forecasting in a building. It should be mentioned that the used dataset for training consists of measured electric loads in an educational building for eight years. In [13], the data have been pre-processed with fussy logic in order to improve the forecasting accuracy and the solar power has been forecasted by ANN. The authors study in [14] presents the forecasting methodology of CO2 emission in Iran in 2030 with using two statistical approaches, multiple linear regression and multiple polynomial regression. Regarding to user convenience and air quality aspect, it is important to get advance knowledge about CO2 concentration to avoid the disadvantages, and to get ready to any change in existing conditions to overcome this issue. This

The present work was done and funded in the scope of the following projects: COLORS Project, CEECIND/02887/2017, and UID/EEA/00760/2019 funded by FEDER Funds through COMPETE program and National Funds through FCT.

978-1-7281-3192-4/19/$31.00 ©2019 IEEE

paper proposes a methodology to forecast CO2 concentration in an office room with using ANN. The used dataset in this paper are the real data which are monitored by existing equipment in the building.

After this section, the methodology is explained in Section II. The case study is demonstrated in Section III and the obtained results will be compared in section IV. Finally, Section VI describes the main conclusions of the work.

II. Methodology Description

The main purpose of this section is to propose the implemented methodology for forecasting the CO2 concentration in buildings. There are several AI techniques for predicting, however ANN is the selected algorithm, since the authors believe that ANN minimizes the error and can provide a predicted value near to actual value. In order to have the precise prediction, it is necessary to use appropriate input data with considering desirable quality and reasonable quantity. These kinds of prediction methodologies are valuable while they are using the real data from the users. Therefore, it is important that all the buildings have been equipped enough in order to monitor the user data to be used in the AI techniques. Fig. 1 shows the overall view of the present methodology.

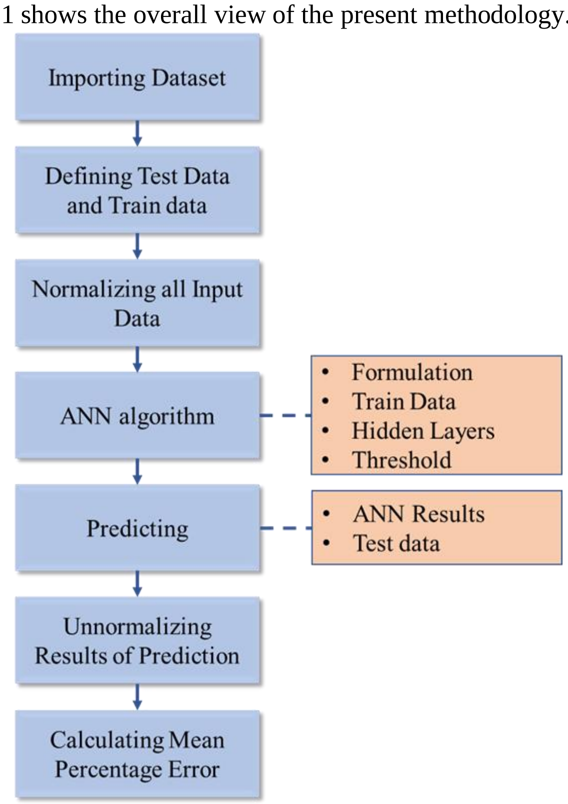

Fig. 1. Flowchart of the methodology.

Input dataset includes train data and test data. The train data consist of input train data and output train data. Output train data can be interpreted as the target of present prediction. It should be noted that precise outcome needs correlated train data. In addition to time and date aspects such as weekday, hour and minute, the historical data of the target for previous moments are also used as input train data. After providing the dataset, all train and test data should be normalized to improve the prediction accuracy.

It can be seen in Fig.1, after defining the required components, the input data should be applied to ANN algorithm. Present methodology is implemented in Rstudio® software with using Neural Net (NN) package which is one of the available packages in Rstudio for implementing forecasting methodologies (www.rstudio.com). There are several optional properties that would be adjusted for predicting in NN function that some of them are mentioned:

- Formulation
- Dataset
- Number of Hidden Layers
- Threshold
- Repetition Times
- Maximum Number of Steps
- Start Weights

More information about NN package is available in (cran.r-project.org). However, according to the purpose, four of them are selected in present algorithm. In addition to train data, a symbolic formulation should be defined to propose the relation of each component of dataset to the target. Equation (1) shows a symbolic example of formulation argument.

$$TrainInput1 + TrainInput2 + \cdots \sim TrainOutput \qquad (1)$$

After defining the formulation, the numbers of hidden layers should be adjusted that it can be any number between numbers of input layer and output layer. The other argument that is used in this study is threshold. A numeric value should be specified for threshold for the partial derivatives of the error function as stopping criteria. After defining all arguments of NN function, the obtained results use in another function to predict the required target. The “Compute” function is available to predict in NN package. Similar to NN function, the compute function contains several arguments with different properties such as:

- NN Results
- Test Data
- Repetition

However, test data and results of NN function are used as arguments in compute function for prediction. Compute function returns a list containing the neuron’s output for each layer of NN. Also, it returns a matrix of overall results of NN. After obtaining the predicted results, they should be unnormalize to compare with initial situation to calculate the mean absolute percentage of error (MAPE) based on (2):

$$MAPE = ABS\left(\frac{Real - Prediction}{Real}\right) \times 100 \qquad (2)$$

MAPE calculation validates the prediction process which is dependent to the select appropriate forecasting algorithm and quality of selected data.

III. Case Study

This section is prepared to test and validate the proposed methodology in a real case study for predicting the CO2 concentration for next day. Therefore, an office in a university building has been selected to perform prediction methodology. The building is located at ISEP Campus, in Porto, Portugal.

This building is equipped with a Supervisory Control And Data Acquisition (SCADA) system for controlling and monitoring several environmental parameters as well as energy consumption and production. In fact, the building is able to manage the electricity consumption and also to record the data in a database. In the SCADA model, there are three distributed based Programmable Logic Controller (PLCs) and one main PLC, which are employed to control and record the data.

Each distributed based PLC is responsible for a group of offices, and the main PLC is responsible to acquire data from the other PLC in order to store in the database, and also to provide a webpage for monitoring the parameters of the building. More detailed information about present SCADA system can be found in [15].

In each office of this building, there are a group of sensors connected to the related PLC. The sensors are CO_2, air quality, humidity, temperature, light intensity, movements and presence of each office users. All these data are transmitted to the PLC through several communication protocols, such as serial and ethernet interface. Also, there are several energy meters in the building that monitor the real-time consumption of lighting system, sockets, and air conditioners.

All these data are acquired from the SCADA model in 1 second time interval, and they are stored in the database with 10 seconds interval. Therefore, the user is able to mine the data from minimum interval of 10 second. This means it is possible to mine the data with bigger time interval for numerous studies.

In this case study, the real data has been acquired from the related database with 1-minute time interval to predict the CO_2 concentration in one specific office.

In order to focus on equal conditions data in terms of daylight and temperature, the used data in this paper is for two periods of winter and autumn seasons:

- October 2017 to March 2018.
- October 2018 to March 2019.

It should be noted that the time aspect has been split to hour and minute in two different input data. The type of used input data in this study are mentioned in below:

- Weekdays
- Hour
- Minute
- CO_2 concentration respectively in 1, 2, 3, 4, and 5 minutes ahead
- CO_2 concentration at corresponding time

In order to increase the quality of input data, only working hours are considered and weekends and holidays and nights hours are ignored. Totally, 242 working days data are imported to the algorithm, which is focused on 10 hours of each day from 10 am to 7 pm. Each hour is split to 60 minutes.

Two different scenarios have been considered for present case study to validate the impact of input data to the prediction accuracy.

It should be mentioned that the only difference of two scenarios is the type of input data and the other parameters are the same. As an example, the quantity of the input data are the same as each other. The purpose of two scenario is forecasting the CO_2 concentration for next day in working hours from 10 am to 7 pm with 1-minute time interval. Table I shows the used input data in each scenario.

TABLE I. USED INPUT DATA FOR EACH SCENARIO

	Train Input				Train Output
Scenarios	Weekdays	Hour	Minute	Historical Data of CO_2	CO_2
Scenario A	×	×	×	×	×
Scenario B	-	-	-	×	×

As it can be seen in Table I, scenario A includes all existing input data in order to predict the CO_2 concentration, while the Scenario B only uses the historical data of CO_2 concentration and has been ignored the date and time data.

It is important to use train input data correlated to target of prediction algorithm. For this purpose, the correlation of all input data to CO_2 amount have been determined and Table II shows the correlation of train input data to the train output. As it can be seen in Table II, the CO_2 concentration in each period is correlated to CO_2 concentration of last periods. In present case, the time interval periods are 1 minute, therefore, according to Table II it can be interpreted that the CO_2 amount is correlated to CO_2 concentration of last 5 minutes.

TABLE II. CORRELATION OF INPUT DATA WITH CO_2 CONCENTRATION.

	Weekday	Hour	Minute	CO2_t-1	CO2_t-2	CO2_t-3	CO2_t-4	CO2_t-5	CO2
Weekday	1								
Hour	-0.00071	1							
Minute	-0.00015	-0.00052	1						
CO2_t-1	-0.03470	0.25716	0.02066	1					
CO2_t-2	-0.03456	0.25800	0.01887	0.99637	1				
CO2_t-3	-0.03442	0.25881	0.01699	0.99224	0.99637	1			
CO2_t-4	-0.03429	0.25959	0.01514	0.98854	0.99224	0.9963769	1		
CO2_t-5	-0.03414	0.26032	0.01352	0.98496	0.98854	0.9922449	0.99637	1	
CO2	-0.03483	0.25629	0.02249	0.99637	0.99224	0.9885415	0.98496	0.98137	1

Also, the correlation numbers of weekdays, hour, minutes show the uncorrelated parameters to the CO2 concentration. However, the next sections will demonstrate the obtained results.Fig.2 presents the CO2 concentration in five different days that are randomly selected from monitored data in the building.

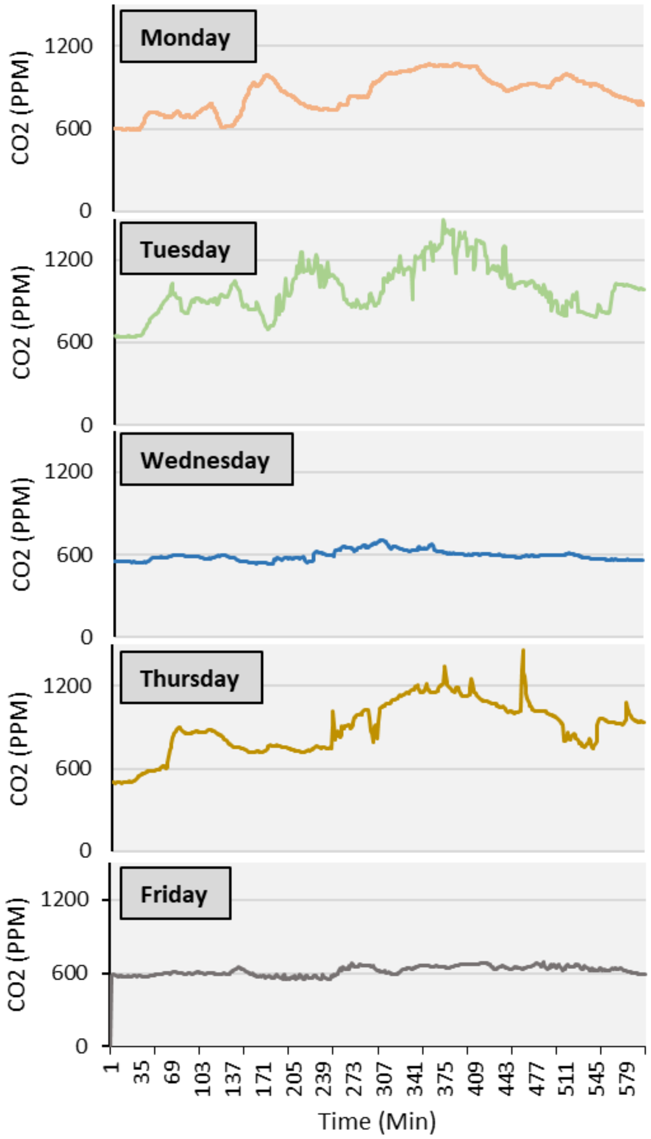

Fig. 2. CO2 concentration in 5 random days.

As it was described in the Section II, there are several available arguments for applying NN function, however, only the train data, formulation, number of hidden layers, and threshold are considered in this case study.

Scenario A

In this scenario, all existing input data such as weekdays, hour, minute and historical data of CO2 concentration in last 5 minutes are applied to input of train data, and the CO2 amount in corresponded time is applied to output or target of train data to the algorithm. The train input data set consists of 242 working days and the algorithm focuses on 10 hours of each day. The According to this train data, the required formulation of the algorithm can be defined. There are 3 hidden layers, and a threshold level equal to 0.1 are considered in this scenario. Test data are related to working hours of 26th March 2019.

Scenario B

In this scenario, the impacts of the date and time on prediction results are studied. For this purpose, the parameters related to the date and times will be ignored, and only the historical data of CO2 concentration in the last minutes would be considered. According to the differences of the train data, the formulation would be also different from the Scenario A. Also, the threshold level and hidden layers in this scenario are considered equal to the Scenario A.

After applying the explained data, the algorithm is executed, and the NN will be trained. After that the gained results with test data will be applied in compute function and the predicted results will be demonstrated in the next section.

IV. Results

The main purpose of this section is to propose the obtained results of two scenarios. Scenario A has been implemented to forecast the CO2 concentration based on date and time information plus historical data of CO2 concentration in last 5 minutes. Fig. 3 shows the predicted CO2 concentration for 600 minutes of next working day. Which is Tuesday, 26th March 2019.

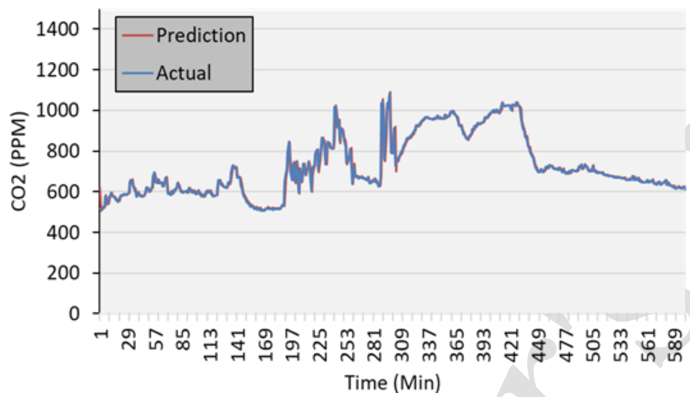

Fig. 3. Results of prediction in scenario A.

As it can be seen in Fig. 3, the predicted CO2 concentration approximately follows the Fig. 2 pattern.

In order to validate the results of prediction, MAPE, Mean Percentage Error (MPE), Standard Deviation (SD), minimum error and maximum error of scenario A are shown in Table III.

TABLE III. Error and Standard Deviation Calculation in Scenario A

Scenario A				
MAPE	MPE	SD	Min MPE	Max MPE
2.044	-0.185	146.207	-27.028	34.343

As it can be seen in Table III, the MAPE value is equal to 2.044 that shows the low level of error in present prediction. However, the SD is equal to 146.204 and can be interpreted that the predicted data are spread out over a wide range.

Fig. 4 presents the overall view of implemented algorithm in the case of network and unnormalized results for scenario A.

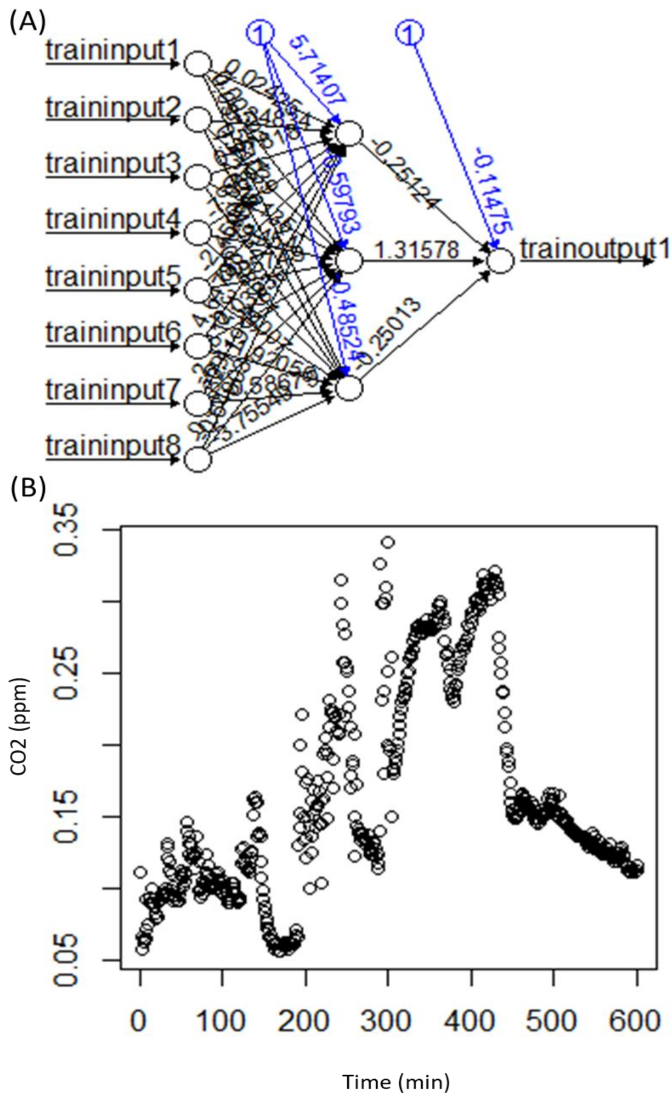

Fig. 4. Overall view of algorithm in scenario A. (A) Diagram of NN. (B)Results of prediction.

Fig. 5 demonstrates the obtained results of scenario B which belongs to 600 minutes of Tuesday 26th March 2019. As it is mentioned in previous section, the predicted data in scenario B is only based on last 5 minutes historical data of CO2 concentration.

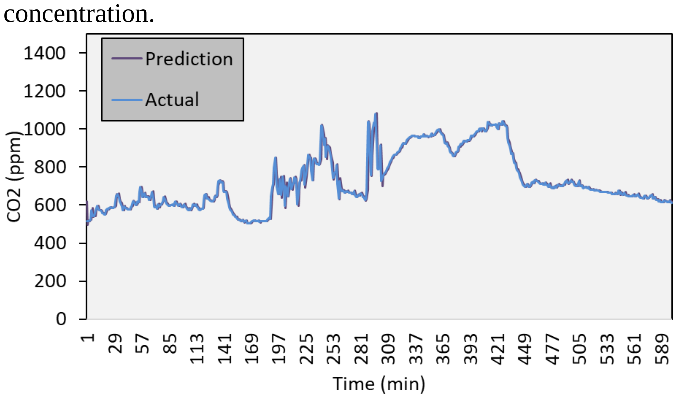

Fig. 5. Results of prediction in scenario B.

As it can be seen in Fig. 5, the actual data has been overlapped the predicted data in scenario B. Some slights variations are visible, but they are not significantly noticeable.

Fig. 6 presents the overall view of implemented algorithm in the case of network and unnormalized results.

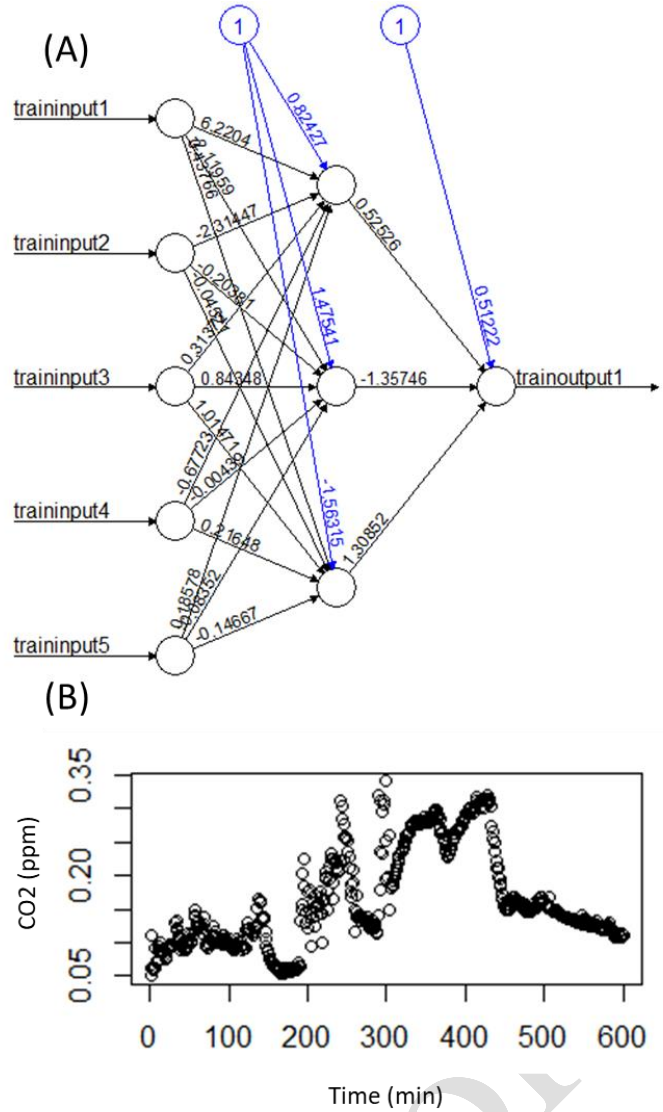

Fig. 6. Overall view of algorithm in scenario B. (A) Diagram of NN. (B)Results of prediction.

Table IV shows the MAPE, MPE, SD, minimum error, and maximum error of gained results of algorithm implementation in scenario B.

TABLE IV. ERROR AND STANDARD DEVIATION CALCULATION IN SCENARIO B

Scenario B				
MAPE	MPE	SD	Min MPE	Max MPE
2.039	-0.15	145.73	-27.11	34.27

According to the comparison of Table III and Table IV it can be seen that two scenarios A and B have approximately the same results.

Fig. 7 presents the comparison of obtained results of two implemented scenarios with actual CO2 concentration.

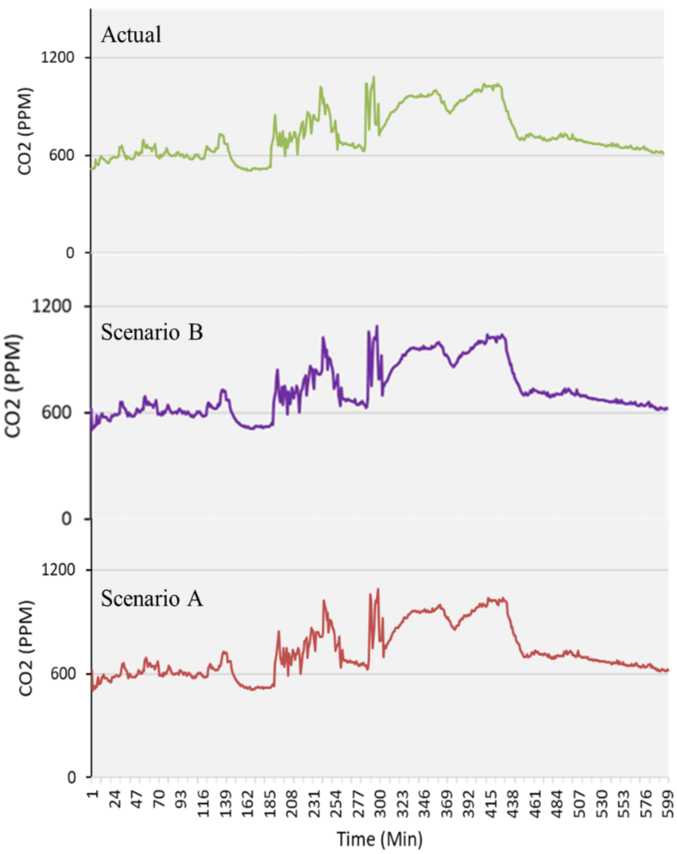

Fig. 7. Comparison of Scenario A and Scenario B with initial case.

According to Fig. 7, it can be seen that two implemented scenarios have been obtained the same results. However, scenario B has been used less input data. In other words, weekdays and time data do not have significant effect on CO2 concentration. Table V shows the comparison of scenario A and scenario B in the field of error percentages and the spread put pattern of result dataset.

TABLE V. COMPARISON OF SCENARIO A AND SCENARIO B

	MAPE	MPE	SD	Min MPE	Max MPE
Scenario A	2.04	-0.18	146.20	-27.02	34.34
Scenario B	2.03	-0.15	145.73	-27.11	34.27

It can be interpreted that scenario B with less numbers of train data have been achieved more desired prediction. Comparing the execution time of two scenarios, the scenario B has been executed significantly rapid with less input data.

V. CONCLUSIONS

This paper presented a forecasting algorithm in order to forecast CO2 concentration in an office room. Artificial neural network as an artificial intelligence technique has been applied to present methodology.

The case study of the present work considered an office building. All the used data had been monitored by several equipment for two years. The train data set selected based on the correlation of monitored data with CO2 concentration in the room. Two different scenarios implemented to survey the impacts of different input data in an equal condition.

The results of paper shown the predicted CO2 concentration, the mean percentage error, standard deviation of result data sets for both situations. This prediction methodology has been used to improve the user convenience. The obtained results of both scenarios presented two equivalent results with different execution time. Equivalent results of two scenario and different execution time illustrated that in some cases a smaller number of input data can increase the efficiency of the algorithm. The obtained results of present algorithm will be used in author's future work in order to control air conditioning system based on CO2 concentration.

REFRENCES

[1] M. Khorram, P. Faria, O. Abrishambaf, and Z. Vale, "Air Conditioner Consumption Optimization in an Office Building Considering User Comfort", International Conference On Energy and Environment Research (ICEER), Aveiro, 2019.

[2] R. Faia, T. Pinto, O. Abrishambaf, F. Fernandes, Z. Vale and J. Corchado, "Case based reasoning with expert system and swarm intelligence to determine energy reduction in buildings energy management", Energy and Buildings, vol. 155, pp. 269-281, 2017.

[3] H. Massrur, T. Niknam and M. Fotuhi-Firuzabad, "Investigation of Carrier Demand Response Uncertainty on Energy Flow of Renewable-Based Integrated Electricity–Gas–Heat Systems", IEEE Transactions on Industrial Informatics, vol. 14, no. 11, pp. 5133-5142, 2018.

[4] A. Jozi, T. Pinto, I. Praça, Z. Vale and J. Soares, "Day ahead electricity consumption forecasting with MOGUL learning model," 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, 2018, pp. 1-6.

[5] F. Silva, B. Teixeira, N. Teixeira, T. Pinto, I. Praca and Z. Vale, "Application of a Hybrid Neural Fuzzy Inference System to Forecast Solar Intensity", 2016 27th International Workshop on Database and Expert Systems Applications (DEXA), 2016.

[6] M. Wen, P. Li, L. Zhang and Y. Chen, "Stock Market Trend Prediction Using High-Order Information of Time Series," in IEEE Access, vol. 7, pp.28299-28308, 2019.

[7] L. Shi and J. Wang, "Application of Artificial Neural Network to Predict the Hourly Cooling Load of an Office Building," 2009 International Joint Conference on Computational Sciences and Optimization, Sanya, Hainan, 2009, pp. 528-530.

[8] H. Wang, Z. Lei, X. Zhang, B. Zhou and J. Peng, "A review of deep learning for renewable energy forecasting", Energy Conversion and Management, vol. 198, p. 111799, 2019. Available: 10.1016/j.enconman.2019.

[9] US EPA. (2019). Carbon Monoxide's Impact on Indoor Air Quality | US EPA. [online] Available at: https://www.epa.gov/indoor-air-quality-iaq/carbon-monoxides-impact-indoor-air-quality.

[10] Occupational Health & Safety. (2019). Carbon Dioxide Detection and Indoor Air Quality Control -- Occupational Health & Safety. [online] Available at: https://ohsonline.com/articles/2016/04/01/carbon-dioxide-detection-and-indoor-air-quality-control.aspx.

[11] Caravaggio, N., Caravella, S., Ishizaka, A. and Resce, G. (2019). Beyond CO2: A multi-criteria analysis of air pollution in Europe. Journal of Cleaner Production, 219, pp.576-586.

[12] L. Xu, S. Wang and R. Tang, "Probabilistic load forecasting for buildings considering weather forecasting uncertainty and uncertain peak load", Applied Energy, vol. 237, pp. 180-195, 2019.

[13] B. Sivaneasan, C. Yu and K. Goh, "Solar Forecasting using ANN with Fuzzy Logic Pre-processing", Energy Procedia, vol. 143, pp. 727-732, 2017.

[14] S. Hosseini, A. Saifoddin, R. Shirmohammadi and A. Aslani, "Forecasting of CO2 emissions in Iran based on time series and regression analysis", Energy Reports, vol. 5, pp. 619-631, 2019.

[15] M. Khorram, O. Abrishambaf, P. Faria and Z. Vale, "Office building participation in demand response programs supported by intelligent lighting management", Energy Informatics, vol. 1, no. 1, 2018.