



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Dataset of short-term prediction of CO₂ concentration based on a wireless sensor network

Ari Wibisono^{a,*}, Hanif Arief Wisesa^a, Novian Habibie^b,
Aulia Arshad^a, Aditya Murdha^a, Wisnu Jatmiko^a, Ahmad Gamal^c,
Indra Hermawan^a, Siti Aminah^a

^a Faculty of Computer Science, Universitas Indonesia, Kampus UI Depok, Indonesia

^b University of Freiburg, Fahrenbergplatz, 79085 Freiburg im Breisgau, Germany

^c Faculty of Engineering, Universitas Indonesia, Kampus UI Depok, Indonesia

ARTICLE INFO

Article history:

Received 16 May 2020

Revised 22 June 2020

Accepted 22 June 2020

Available online 25 June 2020

Keywords:

CO₂ prediction

CO₂ monitoring system

Wireless sensor network

Prediction system

IoT system

ABSTRACT

This CO₂ data is gathered from WSN (Wireless Sensor Network) sensors that is placed in some areas. To make this observation framework run effectively, examining the relationships between factors is required. We can utilize multiple wireless sensor devices. There are three parts of the system, including the sensor device, the sink node device, and the server. We use those devices to acquire data over a three-month period. In terms of the server infrastructure, we utilized an application server, a user interface server, and a database server to store our data. This study built a WSN framework for CO₂ observations. We investigate, analyze, and predict the level of CO₂, and the results have been collected. The Random Forest algorithm achieved a 0.82 R2 Score.

© 2020 The Authors. Published by Elsevier Inc.

This is an open access article under the CC BY license.

(<http://creativecommons.org/licenses/by/4.0/>)

* Corresponding author.

E-mail address: ari.w@cs.ui.ac.id (A. Wibisono).

Specification table

Subject	Computer Science Applications
Specific Subject Area	Computer Science Applications, CO ₂ Prediction
Type of Data	Table Image Chart Graph Figure
How data were acquired	Data is acquired from three sensors. Those sensors are built by authors for the process of data collection. These sensors operated in 24 h and send the data to servers for every second. We have acquired the CO ₂ data for three months period.
Data format	Raw (.csv) comma separated file
Parameters for data collection	The data analysis process is done after the data have been collected over three months. The data acquired for 24 h. There are 6 million instances of sensor readings for the three months. After that, the data are cleared, engineering features are carried out, and models are made to make predictions.
Description of data collection	The data retrieval process is done by placing three nodes in an area. Each node will be connected to 1 other node to be able to send the data from the sensor readings. There is one node sink that is equipped with a GSM to be able to send data to servers that have been online. We have built the hardware and software for the data acquisition process. Data field: <ol style="list-style-type: none"> 1. Record id - Increment record id (unique) 2. Linux timestamp- convert from linux timestamp to time format https://www.epochconverter.com/ 3. CO₂ concentration - in ppm (part per million) 4. Temperature - Temperature in °C 5. Humidity - (Humidity in percentage%) 6. Light intensity - Light Intensity, approx. 0 – 1000 7. Node ID Node 0, Node 1, and Node 9 <p>The data generated from the sensors can be used to form a model for measuring and predicting future CO₂ measurements. Because the time interval for recording CO₂ level is very short, many data are generated. This makes it very possible to predict CO₂ values based on time. This method can be very feasibly implemented in various places</p> <p>The data generated are approximately 6 million instances. The prediction is done by dividing the data into 2 parts, namely, 80% of the data are used for training the machine learning models and 20% of the data are used for testing. We do 10-fold cross-validation to measure the performance of each algorithm against the sensor acquisition data. The metrics used are standard metrics, namely, the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the R2 Score.</p> <p>Fig. 1, Fig. 5, Fig. 6, Fig. 7 are our framework design to gathered environment data. Fig. 2 shows the frequency distribution of each feature. In Fig. 2. (a) is the graph for the CO₂ data. The most frequent value for CO₂ data is approximately 200 - 300 ppm. As we know, the global CO₂ concentration is approximately 450 ppm. The measurement results show that the CO₂ concentration is are lower than the global condition. Fig. 2. (b) shows the results of the air temperature measurements. In general, the most frequent value is between 25 and 30 °Celsius. Sometimes the temperature can drop to below 20° when the weather is cold and increase to more than 35° during the day. These data are taken in a tropical environment, which means that there is no winter, summer, autumn, or spring condition.</p> <p>Based on the observations in Table 1, there are two directions, positive and negative correlations. These are the conclusions of the correlation analysis of the data. Regarding the correlation of CO₂ - Air Humidity, from the analysis of the gathered data, the correlation between CO₂ and humidity is consistently more than 0.6. The mean correlation is 0.63. A strong correlation with a positive direction is picturized by the Cohen Scale. This means that the two parameters move in the same direction. According to Fiddbruk, the correlation</p>

(continued on next page)

Data source location	<p>between CO₂ and humidity happens in indoor situations since humans exhale CO₂, which produces water vapor [7]. This may also be true for the outdoor situation since there are many organisms and machinery that emit CO₂. As can be seen in Fig. 8, the MAE measurement results show that a fairly low error is obtained by the Random Forest (RF) algorithm with an average MAE of 37.07 and that for the Gradient Boosting Regressor (GBR) was 36.77. High errors are obtained by the Linear Regression (LR) and Ridge Regression (RR) at 63.53. The Decision Tree Regressor (DT) has a fairly small error of 45.3. be seen in Fig. 9, the RMSE measurement results show that a fairly low error is obtained by the Random Forest (RF) algorithm with an average RMSE of 51.2 and that for the gradient Boosting Regressor (GBR) was 49.06. High errors are obtained by the Linear Regression (LR) and Ridge Regression (RR) at 91.9. The Decision Tree Regressor (DT) has a fairly small error of 63.32. To assess the quality of the prediction results, we evaluate the R2 score, which is in the range of 0 - 1. The measurement results of the 10-fold cross validation are described in the form of boxplots, where one boxplot can represent a general measurement result, the highest measurement results, the lowest measurement results, and the median. As can be seen in Fig. 10, the measurement results of the R2 Score show that a high enough score is obtained by the Random Forest (RF) algorithm with an average R2 Score of 0.78 and that for the Gradient Boosting Regressor (GBR) was 0.82. Low R2 Score results are obtained by the Linear Regression (LR) and Ridge Regression (RR) at 0.43. The Decision Tree Regressor (DT) has a fairly large R2 of 0.72. All of the dataset are available in Mendeley data repository. Institution: Faculty of Computer Science Universitas Indonesia City/Town/Region: Depok/West Java Country:Indonesia</p>
Data accessibility	<p>Upload to mendeley https://data.mendeley.com/datasets/6d798dkhpz/draft?a=502e7fa1-e1b4-4931-99eb-1aeb4ecc4f13 Source code: https://github.com/WSN-1231/</p>
Related research article	<p>Habibie et al., "Comparative study of lightweight secure multiroute communication system in low cost wireless sensor network for CO₂ monitoring," 2016 International Workshop on Big Data and Information Security (IW BIS), Jakarta, 2016, pp. 145–150, doi: 10.1109/IWBIS.2016.7872904. A. Arshad, N. Habibie, A. Wibisono, P. Mursanto, W. S. Nugroho and W. Jatmiko, "Sensor node for data sampling and correlation analysis of CO₂ concentration with air humidity, temperature, and light intensity," 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Malang, 2016, pp. 111–116, doi: 10.1109/ICACSIS.2016.7872793.</p>

Value of data

- This data consists of the results of environmental sensor readings which useful to support monitoring and prediction of environmental changes. This data is important because it gathered from several sensors regarding environmental conditions. The parameters measured are temperature, CO₂ concentration, humidity, and light intensity.
- All institutions, governments, and private sectors that have intention on monitoring and predicting environmental conditions will benefit from this data.
- The added value of this data is the fast time-taking interval of the data gathering process. The sensors gathered the environment condition every 1 s for 24 h.
- A framework to gathered the environment data has been proposed in this data. It is implemented using WSN to observe CO₂ concentration, temperature, light intensity, and humidity. Framework integration between hardware devices, protocols, wireless sensor networks, and servers.
- This data is a representation for evaluating and predicting rapid environmental changes within a certain period. Data Analysis correlation factors that influence the measurement of CO₂ concentration

- In short-term use, this data will be useful for recognizing the rapidly changing environmental conditions of an area. In the long term, this data can recognize local changes that will shape the behavior of environmental conditions over a long period.

1. Data

The data retrieval process is done by placing three nodes in an area. Each node will be connected to 1 other node to be able to send the data from the sensor readings. There is one node sink that is equipped with a GSM to be able to send data to servers that have been online. Each sensor is equipped with a 12V 65A battery. The battery's ability to supply power to the equipment is five days. Every five days, the battery is replaced for each node.

The most feasible method to gather the data and monitor the data is by using a Wireless Sensor Network (WSN). The main function of a WSN is to monitor several remote [1,2]. A WSN provides a framework using a network of sensors to observe those areas. It can primarily be used to monitor natural resources, telecommunications, and others. There has been much research that was conducted related to Wireless Sensor Networks (WSNs). Oihha et al. used a WSN to monitor farming [2] and Rashid et al. used the framework for urban territory monitoring. Mekki et al. used a WSN to monitor the CO₂, temperature, and dampness light to observe the natural environment [3]. Another research conducted by Sudarrono examined secure data transmission using a WSN by using light encryption that is implemented inside the WSN framework [4-6].

The main advantage of using the WSN framework is its size and cost. The sensors in this framework are very compact and affordable. Additionally, it is very efficient in terms of its power consumption. This makes it more feasible for it to be implemented in remote areas where remote monitoring will be done using microcontrollers. However, its computational and memory power are limited due to the low-end grade of the controller. Therefore, it is a challenge to implement many available algorithms due to the limitation of this computational resource. A correct and suitable algorithm needs to be implemented in the WSN.

1.1. Data acquisition process

Throughout the span of three months, we obtained 6 million instances to be used to analyze and model CO₂ behavior. The Data Logging and sensor nodes' data acquisition occur every 1 s.

Fig. 1 shows the frequency distribution of each feature. In Fig. 2(a) is the graph for the CO₂ data. The most frequent value for CO₂ data is approximately 200 - 300 ppm. As we know, the global CO₂ concentration is approximately 450 ppm. The measurement results show that the CO₂ concentration is are lower than the global condition. Fig. 2(b) shows the results of the air temperature measurements. In general, the most frequent value is between 25 and 30 °Celsius.

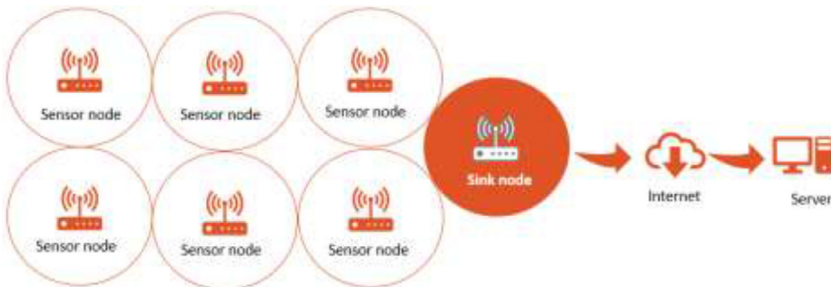


Fig. 1. Design of the WSN architecture.

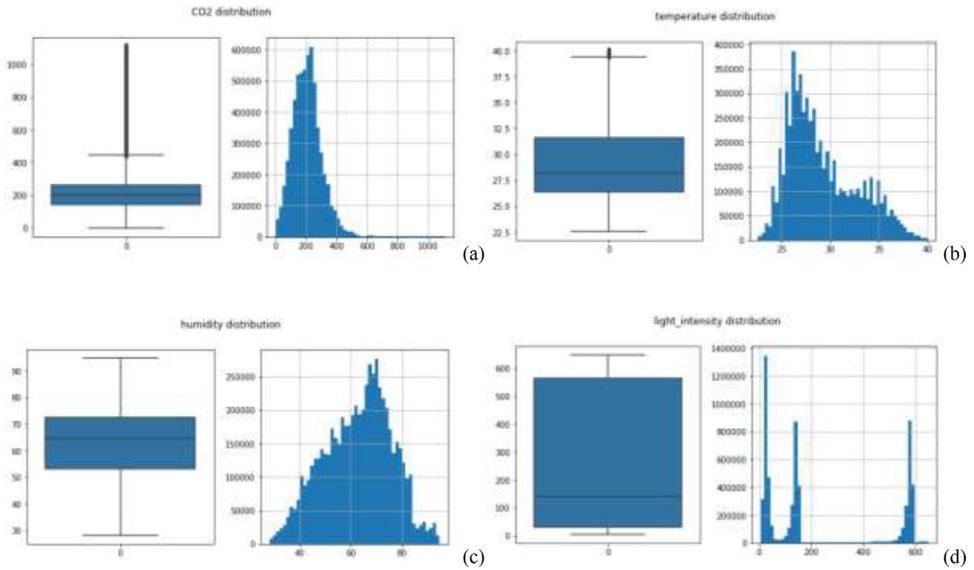


Fig. 2. Distribution frequencies of the data features: (a) CO₂, (b) temperature, (c) humidity, and (d) light intensity.

Sometimes the temperature can drop to below 20° when the weather is cold and increase to more than 35° during the day. These data are taken in a tropical environment, which means that there is no winter, summer, autumn, or spring condition.

Fig. 2(c) shows the frequency distribution for the air humidity data. The most frequent value is between 60 - 70%. As we know, the tropics usually have a high level of humidity. Fig. 2(d) shows the frequency distribution of the light intensity. In the frequency distribution in Fig. 2(d), there are two significant peak frequencies, among others. First is the intensity of the light between 0 and 200, which is the change in the intensity from the evening - night and morning to noon. The next frequency distribution is between 500 - 600, which is the light intensity during the day.

1.2. Data analysis

The data analysis process is done after the data have been collected over three months. There are 6 million instances of sensor readings for the three months. After that, the data are cleared, engineering features are carried out, and models are made to make predictions.

Fig. 3 shows a description of the correlation of each feature with the CO₂ variable. It can be seen in Fig. 3(a) that there is a decreasing trend between the temperature and CO₂ features. This shows a strong negative correlation between the two features. From the results of the measurement of the correlation score between the two variables in Fig. 4, the two variables have a high correlation score of 0.63. In Fig. 3(b), (c) and (d), there is no visible up or downtrend with respect to the CO₂ variable such that the correlations given in Fig. 8 is not too large between these variables with respect to CO₂, which are below 0.2.

The correlation analysis was conducted using the 24 h measurements. Each day may have different data characteristics. Therefore, we change the analysis to daily analysis. An example of the reason for this is that the CO₂ concentration during working days is relatively higher than the concentration during the weekend.

Based on the observations in Table 1, there are two directions, positive and negative correlations. These are the conclusions of the correlation analysis of the data. Regarding the correlation of CO₂ - Air Humidity, from the analysis of the gathered data, the correlation between CO₂ and

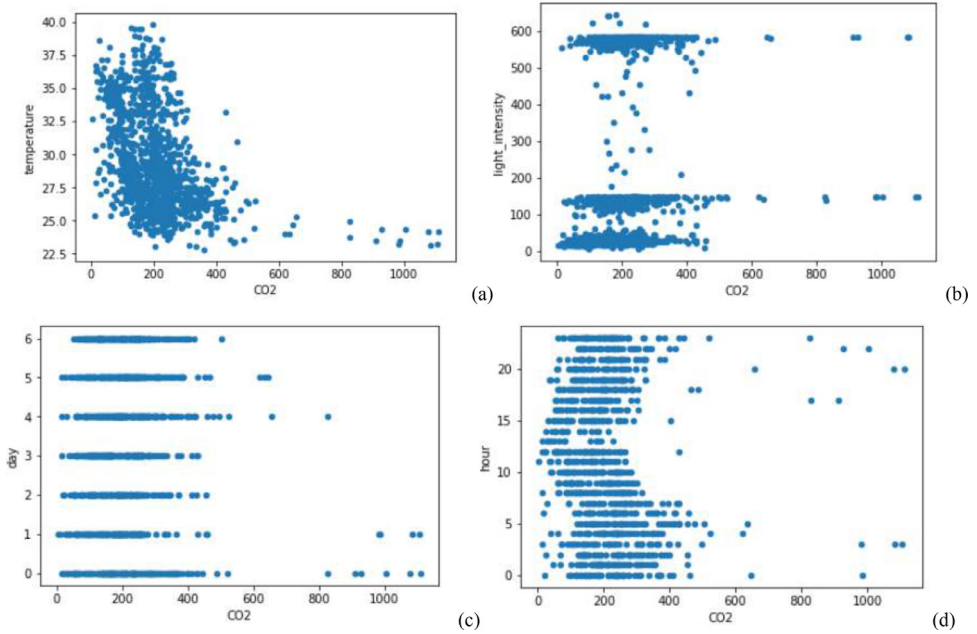


Fig. 3. Correlation of (a) Temperature, (b) Light Intensity, (c) day, and (d) hour with CO₂.

	CO2	temperature	humidity	light_intensity	node_id	hour	minute	day
CO2	1.0	-0.44	0.63	0.19	-0.18	-0.17	-0.011	0.021
temperature	-0.44	1.0	-0.87	-0.48	0.0042	0.27	-0.0038	0.069
humidity	0.63	-0.87	1.0	0.41	-0.048	-0.27	-0.0012	0.01
light_intensity	0.19	-0.48	0.41	1.0	0.029	0.024	0.0092	-0.017
node_id	-0.18	0.0042	-0.048	0.029	1.0	-0.0069	0.0071	-0.0082
hour	-0.17	0.27	-0.27	0.024	-0.0069	1.0	-0.0018	-0.0089
minute	-0.011	-0.0038	-0.0012	0.0092	0.0071	-0.0018	1.0	-0.0028
day	0.021	0.069	0.01	-0.017	-0.0082	-0.0089	-0.0028	1.0

Fig. 4. Pearson correlation scores of features.

Table 1
Correlation between light power, air moistness, and air temperature with CO₂.

Parameters	Correlation value	Result
CO ₂ + Humidity	0.63	+ Strong correlation (A)
CO ₂ + Temperature	- 0.44	- normal Correlation (B)
CO ₂ + Light Intensity	0.19	+ weak correlation (C)
CO ₂ + Hour	-0.17	- weak correlation (C)
CO ₂ + Minute	-0.011	- weak correlation (D)
CO ₂ + Day	0.021	+ weak correlation (D)

humidity is consistently more than 0.6. The mean correlation is 0.63. A strong correlation with a positive direction is picturized by the Cohen Scale. This means that the two parameters move in the same direction. According to Fiddbruk, the correlation between CO₂ and humidity happens in indoor situations since humans exhale CO₂, which produces water vapor [7]. This may also be true for the outdoor situation since there are many organisms and machinery that emit CO₂.

One of the researchers concludes that the air temperature is not correlated with the CO₂ concentration, where the CO₂ does not contribute to temperature changes [8]. However, in this dataset, regarding the correlation of CO₂ - Air Temperature, there is a consistent correlation of CO₂ and temperature. It resulted in a Pearson Correlation Score of -0.44. Based on our analysis, the correlation between the light intensity and CO₂ concentration is weak. It is also negative. From the data samples, almost all of them have weak correlations. This condition occurs due to the effect of photosynthesis in plants, where the higher light intensity will reduce the CO₂ intensity surrounding the plants (since the photosynthesis process uses CO₂) [9,10]. However, the results may vary in open air since the original work was done in a closed room.

The time parameter is also considered in this analysis. The results were -0.17, -0.011, and 0.021 for CO₂ + Hour, CO₂ + Minute, and CO₂ + Day, respectively. Although the correlation score of the time parameter is below 0,2, we still need to consider these parameters as features for the machine learning evaluation since these time parameters influence the temperature and light intensity. The temperature, light intensity, and humidity are influenced by these time parameters.

2. Experimental, system design, and prediction methods

Carbon dioxide (CO₂) is a gas that exists in our atmosphere. In the correct amount, it could balance the atmosphere and create a more stable condition. In recent years, the concentration of CO₂ around the world has increased very rapidly with an increase of 21.12 ppm from 2006 to 2015 [11]. The vast increase in the CO₂ concentration in our atmosphere could cause an imbalanced air composition, which could lead to several chronic diseases. These diseases include respiratory issues, vision issues, and other complications [12]. Additionally, the rise in the CO₂ concentration could also cause an increase in the global temperature. The agricultural sector could also be affected by the increase in the CO₂ concentration since an excessive amount of CO₂ being processed during the photosynthesis of plants could actually reduce the nutrients contained in the crops [13,14].

To tackle this issue, more detailed monitoring of the CO₂ concentration is required. The data gathered from the monitoring process will be used for decision making in order to hamper the increase in the CO₂ concentration. To do this, we need to have a state-of-the-art CO₂ monitoring system. A feasible solution is using a Wireless Sensor Network (WSN) to acquire these CO₂ concentration data. The sensors are placed in certain locations to gather data, which are then shared with the other sensors in the network to analyze the CO₂ data.

To assist the data analysis process, other additional parameters are required so that the exact state is known in detail. These parameters include the temperature, air humidity, and light intensity. However, not all of these parameters will contribute substantially to the analytical process. Therefore, we need to select the parameters that will help the CO₂ data analysis process. To do this, a correlation evaluation between the parameters and the CO₂ concentration is needed.

Various research has been done in order to examine the correlation of those parameters with the CO₂ concentration. Li, et al. analyzed the concentration of CO₂ in different altitudes in the atmosphere [15]. To do this, the authors attached a CO₂ sensor on an air balloon. The results showed that the CO₂ concentration is lower at the higher altitudes of our atmosphere. This means that the altitude is inversely correlated with the CO₂ concentration. Another study in 2013 analyzed the correlation of the CO₂ concentration with another parameter. Katarzyna et al. analyzed the correlation between the indoor CO₂ concentration and air humidity [7]. Lazovic et al. also studied the correlation of the humidity and air temperature [16]. Another correlation study conducted by Soares analyzed the correlation between CO₂ and the temperature in order

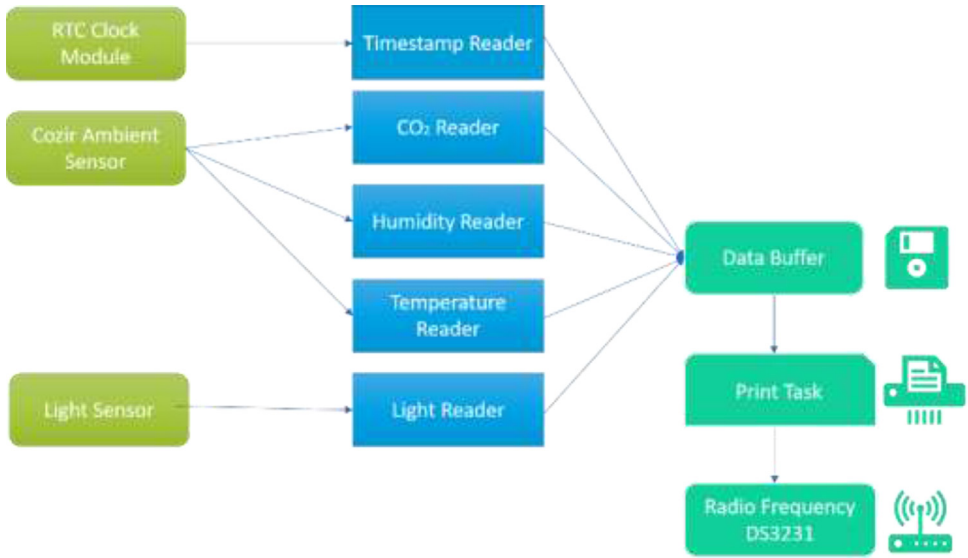


Fig. 5. RTOS configuration.

to analyze global climate change. A. Arshad, et al. conducted data sampling and a correlation test for CO₂ [17].

2.1. System design

2.1.1. Hardware design

There are a few parts in our hardware such as the timer RTC module, the Radio Frequency (RF) communication module sensor, the light intensity sensor, the integrated environmental sensor (CO₂ sensor, temperature sensor, and humidity sensor), and the 256 KB Microcontroller. We utilize a Single Board Computer (SBC) for temporary storage. The microcontroller controls the various sensors that transmit the captured data and stores the captured data in a Raspberry Pi 2 (Single Board Computer). After the data from many nodes are stored in our SBC, we send all captured data to our server using a 4G connection via the internet.

2.1.2. Software design

The Raspberry Pi 2 has a Linux Raspbian Operating System. Qt5 is used to develop the temporary storage, which is compiled in the Raspberry Pi 2. The temporary storage is connected to the Microcontroller. The 256 KB Microcontroller uses the Real Time Operating System (RTOS). The RTOS can perform multitasking, which is divided into tasks independent of each other, and is also better than the primitive interrupt for an embedded system [18]. The process inside the sensor node is divided into five tasks: the retrieval of CO₂, temperature, humidity, and light intensity data and integration. The RTOS that is used in this study is FreeRTOS [19]. The configuration can be found in Fig. 5.

2.1.3. WSN design

In this research, the WSN system will also be integrated with the server. As mentioned in the previous section, the input data will be gathered by the sensor nodes, which are installed according to the testing. The data that were gathered by these nodes will be stored inside the centralized storage. The temporary storage will act as the producer that transmits the data to

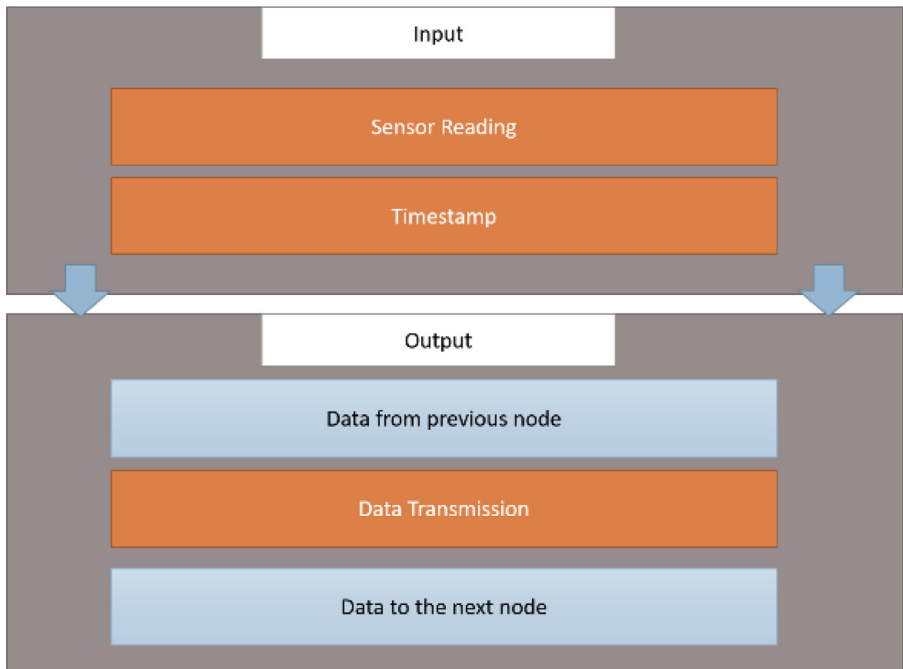


Fig. 6. The data transmission flow.

the server web service. The server is the consumer, which will gather the data and relay it to the SQL database. Fig. 6. shows the information flow in the sensor node.

Multiple nodes will establish a radio-based communication network in the zone. There are two types of WSN nodes: the source node and the sink node. The source node acquires data from the sensor, and the sink node is responsible for transmitting the data acquired by the source node to the server. These two processes should be done simultaneously during their execution. Sensor tasks are implemented in the Microcontroller by using the RTOS. The details of the RTOS task flow in the Microcontroller are described in Fig. 7.

2.2. Experiment result and prediction system

This experiment obtains time series data to analyze the correlation of one parameter with the other parameters. The sampling has been done using the proposed WSN nodes in a specific area.

In the experiment, the sensor nodes were placed on a road with 1.5 m between the nodes and the road. Data logging was conducted over two weeks. According to Li, the CO₂ sampling that is conducted at a very low height will result in a higher concentration of CO₂ [15]. Therefore, we have placed the nodes as low as possible, which is 1 m due to safety reasons.

The data generated from the sensors can be used to form a model for measuring and predicting future CO₂ measurements. Because the time interval for recording CO₂ level is very short, many data are generated. This makes it very possible to predict CO₂ values based on time. This method can be very feasibly implemented in various places.

The data generated are approximately 6 million instances. The prediction is done by dividing the data into 2 parts, namely, 80% of the data are used for training the machine learning models and 20% of the data are used for testing. We do 10-fold cross-validation to measure the performance of each algorithm against the sensor acquisition data. The metrics used are standard

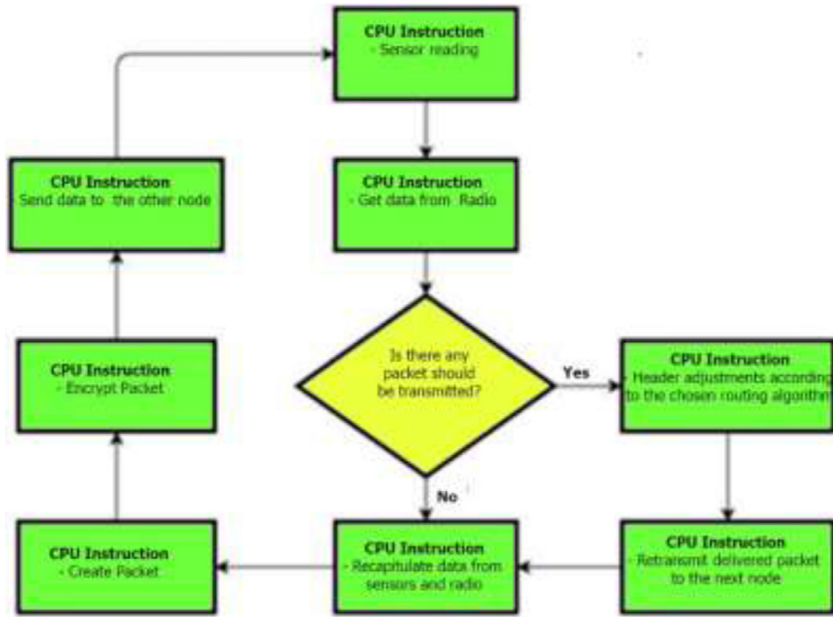


Fig. 7. RTOS tasks in the microcontroller.

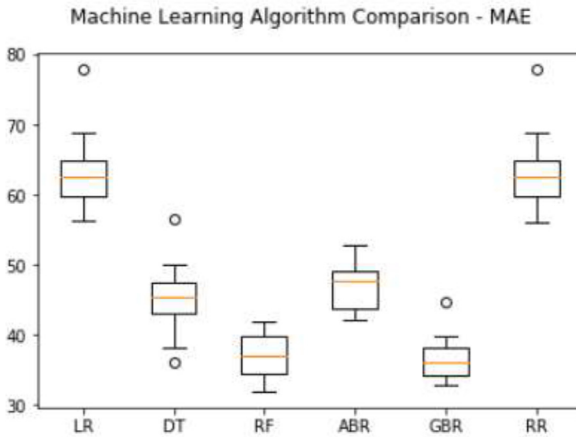


Fig. 8. MAE comparison result.

metrics, namely, the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE), and the R2 Score.

The measurement results of the 10-fold cross validation are described in the form of boxplots, where one boxplot can represent a general measurement result, the highest measurement results, the lowest measurement results, and the median. As can be seen in Fig. 8, the MAE measurement results show that a fairly low error is obtained by the Random Forest (RF) algorithm with an average MAE of 37.07 and that for the Gradient Boosting Regressor (GBR) was 36.77. High errors are obtained by the Linear Regression (LR) and Ridge Regression (RR) at 63.53. The Decision Tree Regressor (DT) has a fairly small error of 45.3.

The measurement results of the 10-fold cross validation are described in the form of boxplots, where one boxplot can represent a general measurement result, the highest measurement

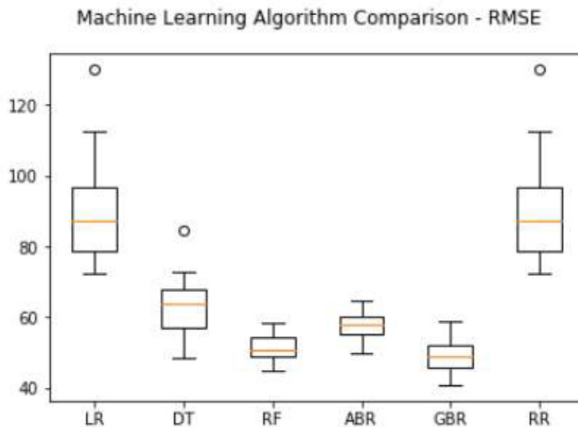


Fig. 9. RMSE comparison result.

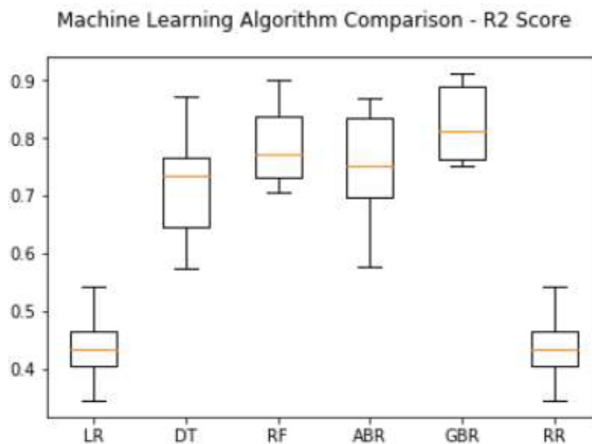


Fig. 10. R2 score comparison result.

results, the lowest measurement results, and the median. As can be seen in Fig. 9, the RMSE measurement results show that a fairly low error is obtained by the Random Forest (RF) algorithm with an average RMSE of 51.2 and that for the gradient Boosting Regressor (GBR) was 49.06. High errors are obtained by the Linear Regression (LR) and Ridge Regression (RR) at 91.9. The Decision Tree Regressor (DT) has a fairly small error of 63.32.

To assess the quality of the prediction results, we evaluate the R2 score, which is in the range of 0 - 1. The measurement results of the 10-fold cross validation are described in the form of boxplots, where one boxplot can represent a general measurement result, the highest measurement results, the lowest measurement results, and the median. As can be seen in Fig. 10, the measurement results of the R2 Score show that a high enough score is obtained by the Random Forest (RF) algorithm with an average R2 Score of 0.78 and that for the Gradient Boosting Regressor (GBR) was 0.82. Low R2 Score results are obtained by the Linear Regression (LR) and Ridge Regression (RR) at 0.43. The Decision Tree Regressor (DT) has a fairly large R2 of 0.72.

In addition to using the classical machine learning algorithm to model CO₂, we also use the RNN (Recurrent Neural Network) method. The reason for using the RNN method is because this algorithm is a derivative of deep learning that uses a recurrent network mechanism to handle time series data. It is compatible with the CO₂ time series dataset. The RNN uses a mechanism

to connect the previous state data ($t-1$) to the next state (t), and so it considers the previous time state ($t-1$).

The mechanism of LSTM (Long Short Term Memory) is used in the RNN to model the short term prediction of CO₂ data. The results we get from the RNN simulation produce an MAE of 39.89 and an RMSE of 56.34. We did not directly compare the results of the RNN method and those of the classical machine learning method because the RNN training was done 50 times and deep learning uses iterative learning for each training step that has been conducted. However, for classical machine learning, training is only done once, and this is enough since it is not iterative learning.

It can be seen from the results of the MAE, RMSE, and R2 Score that the best algorithm that obtains the least error and the highest score is the Gradient Boosting Regression (GBR) with an R2 Score of 0.82, an MAE of 37.07, and an RMSE of 49.06. Therefore, we can conclude that the prediction system using the GBR algorithm is quite good at making short-term predictions of the CO₂ concentration.

A WSN is a framework that produces huge amounts of information. Hence, a framework that can competently deal with the enormous amounts of data is required. As a solution, we proposed a framework model that can be used as a tool to measure and predict the CO₂ concentration. The collected data are analyzed by the machine learning algorithm that has achieved excellent accuracy performance. It results in an R2 score of 0.82. In further research, we are going to analyze CO₂ data combined with another dataset, e.g., CCTV data, to predict the CO₂ concentration.

Authors' contributions

- Ari Wibisono: Managed the technical implementation and the testing scenario, and combined all the technical results.
- Hanif Arief Wisesa: Constructed and developed the technical materials of this paper.
- Wisnu Jatmiko: Head of the project, and wrote the research background and introduction
- Ahmad Gamal: Vice head of the project, and wrote the research background and introduction
- Siti Aminah and Aulia Arshad: Conducted the statistical analyses and wrote the results
- Novian Habibie: Integrated and implemented the application and WSN
- Aulia Arshad: Applied, implemented and evaluated the sensor nodes
- Indra Hermawan: Conducted the device configuration and data acquisition
- Aditya Murda: Conducted the WSN implementation and evaluation

Declaration of Competing Interest

The authors declare that they have no competing interests.

Acknowledgments

This article's publication is partially supported by the [United States Agency for International Development](#) (USAID) through the Sustainable Higher Education Research Alliance (SHERA) Program for Universitas Indonesia's Scientific Modeling, Application, Research, and Training for City-centered Innovation and Technology (SMART CITY) Project, Grant #AID-497-A-1600004, Sub Grant#IIE- 00000078-UI-1.] No:0141 /UN2.R3.SC/ HKP. 05.01/2018. We affirm that the submission represents original work that has not been published previously and is not currently being considered by another journal. Additionally, we confirm that each author has seen and approved the contents of the submitted manuscript. This work is supported by the Faculty of Computer Science, [Universitas Indonesia](#)

Declarations

Availability of data and material: All material and source codes are available at this link: <https://github.com/WSN-1231/>.

Funding: United States Agency for International Development (USAID) through the Sustainable Higher Education Research Alliance (SHERA) Program for Universitas Indonesia's Scientific Modeling, Application, Research and Training for City-centered Innovation and Technology (SMART CITY) Project, Grant #AID-497-A-1600004, Sub Grant#IIE- 00000078-UI-1.] No:0141 /UN2.R3.SC/HKP. 05.01/2018

Ethics

This work didn't involve human or animal subjects.

References

- [1] B. Rashid, M.H. Rehmani, Applications of wireless sensor networks for urban areas: a survey, *J. Netw. Comput. Appl.* 60 (2016) 192–219.
- [2] T. Ojha, S. Misra, N.S. Raghuvansfa, Wireless sensor networks for agriculture: The rtatd-bf-thd-art in practice and future challenges, *Comput. Electron. Agric.* 118 (2015) 66–84.
- [3] M. Mekki, O. Abdallah, M.B.M. Amin, M. Ebaye, R.T. Abdalfatah, A. Babiker, Greenhouse monitoring and control system based on wireless sensor network, in: *Proceedings of the International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering (ICNEEEE)*, Sept 2015, pp. 384–387.
- [4] A. Sudarsono, P. Kristalina, M.U.H. Al Rasyid, R. Hermawan, An implementation of secure data sensor transmission in wireless sensor network for monitoring environmental health, in: *2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA)*, Bandung, 2015, pp. 93–98, doi:10.1109/IC3INA.2015.7377753.
- [5] J.-A. Jiang, C.-H. Wang, C.-H. Chen, M.-S. Liao, Y.-L. Su, W.-S. Chen, C.-P. Huang, E.-C. Yang, C.-L. Chuang, A wsn-based automatic monitoring system for the foraging behavior of honey bees and environmental factors of beehives, *Comput. Electron. Agric.* 123 (2016) 304–318.
- [6] F.M. Al-Turjman, H.S. Hassanein, M. Ibnkahla, Towards prolonged lifetime for deployed (WSNs) in outdoor environment monitoring, *Ad Hoc Netw.* 24 (2015) 172–185 Part A.
- [7] K. Gadyrzdwrka-Fiddbruk, Correlation of air humidity and carbon dioxide concentration in the kindergarten, *Energy Build.* 62 (2013) 45–50.
- [8] P.C. Soares, et al., Warming power of CO₂ and H₂O: correlations with temperature changes, *Int. J. Geosci.* 1 (03) (2010) 102.
- [9] M. Martin, D.M. Noarbe, P.H. Serrot, B. Sabaler, The rise of the photo-synthetic rate when light intensity increases is delayed in *ndh* gene- defective tobacco at high but not at low CO₂ (bn(dnratibnr, *Front. Plant Sci.* 6 (2015) 34.
- [10] T. Raki, G. Gaji, M. Lazarevi, B. Stevanovi, Effects of different light intensities, {CO₂} concentrations, temperatures and drought stress on photobrynthetis(activity in two paleoendeiTiic resurrection plant species ramonda serbica and r. nathaliae, *Environ. Exp. Bot.* 109 (2015) 63–72.
- [11] E.S.R. Laboratory. Trends in atmospheric carbon dioxide. [Online]. Available: <https://www.esrl.noaa.gov/gmd/ccgg/trends/history.html>
- [12] Guan, Wei-Jie, et al., Impact of air pollution on the burden of chronic respiratory diseases in China: time for urgent action, *Lancet* 388. 10054 (2016) 1939–1951. L. H. Dietterich and et al., Impacts of elevated atmospheric CO₂ on nutrient content of important food crops, *Sci. Data* 2 (2015) 150036.
- [13] S.S. Myers, et al., Increasing CO₂ threatens human nutrition, *Nature* 510 (2014) 139–142.
- [14] Y. Li, J. Deng, C. Mu, Z. Xing, K. Du, Vertical distribution of CO₂ in the atmospheric boundary layer: Characteristics and impact of meteorological variables, *Atmos. Environ.* 91 (2014) 110–117.
- [15] I.M. Lazovic, Z.M.V. Stevanovic, M.V. Jovasevic-Stojanovic, M.J.B. Zivkovic, Impact of CO₂ concentration on indoor air quality and correlation with relative humidity and indoor-air temperature in school buildings, *Therm. Sci.* 20 (00) (January 2015) 173–173.
- [16] Arshad, N. Habibie, A. Wibisono, P. Mursanto, W.S. Nugroho, W. Jatmiko, Sensor node for data sampling and correlation analysis of CO₂ concentration with air humidity, temperature, and light intensity, in: *Proceedings of the International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, Malang, 2016, pp. 111–116. 2016.
- [17] J. Norman, An aodv Network Layer for nrf4101+ (or similar) Chips for Arduino, July 2015.
- [18] N. Habibie, et al., Comparative study of lightweight secure multiroute communication system in low cost wireless sensor network for CO₂ monitoring, in: *Proceedings of the 2016 International Workshop on Big Data and Information Security (IWBSIS)*, Jakarta, 2016, pp. 145–150. 2016.