Computer Engineering and Applications Vol. 10, No. 1, February 2021

Air Quality Classification Using Support Vector Machine

Ade Silvia Handayani[1,*], Sopian Soim[1], Theresia Enim Agusdi[1], Nyayu Latifah Husni[2]

[1]*Department of Telecommunication Engineering, State Polytechnic of Sriwijaya*
[2]*Department of Electrical Engineering, State Polytechnic of Sriwijaya*
**ade_silvia@polsri.ac.id*

ABSTRACT

Air pollution in Indonesia, especially in urban areas, becomes a serious problem that needs attention. The air pollution will impact on the environment and health. In this research, the air quality will be classified using Support Vector Machine method that obtained from the sensor readings. The sensors used in the detection of CO, CO_2, HC, dust/PM_{10} and temperature, namely TGS-2442, TGS-2611, MG-811, GP2Y1010AU0F and DHT-11. After testing, the results obtained with classification accuracy of 95.02%. The conclusion of this research indicates that the classification using the Support Vector Machine has the ability to classify air quality data.

Keywords: Air Quality, Artificial Intelligence, Machine Learning, Classification, Support Vector Machine.

1. INTRODUCTION

Air pollution is an event in which air pollutant gaseous pollutants replace the function of air, so that it can disrupt daily activities [1]. As for some very dangerous pollutants such as carbon, NO_2, SO_2, particulate matter (PM) can affect the quality of life, health problems, and the environment [2][3]. Indonesia is one of the most polluted cities. This is evidenced by the 30% increase in vehicle growth, especially motorbikes [4]. The large number of vehicles results in high levels of air pollutants which have a negative impact on public health such as acute respiratory infections, bronchitis, skin irritations, and cardiovascular disease [5]. In 2015, 81% of the total emissions from forest and land fires that occurred in Indonesia resulted in greenhouse gas emissions [6]. Other than that, temperature changes caused by climate change directly or indirectly have a serious impact on human disease mortality or morbidity [7].

The Indonesian government has made various efforts in overcoming air pollution, including by issuing Peraturan Pemerintah No. 41 in 1999 concerning air pollution control [8]. Several research about control and prevention of air pollution have been done such as expanding green infrastructure, emphasizing multi-pollutant emissions reductions, and implementing car free days [9][10][11]. Meanwhile, the implementation of Autonomus Vehicle (AV) in research [12] has purpose to analyze the potential of AV in terms of reducing fuel consumption and greenhouse gas emissions from transportation, both technically and technically. However, various existing policies and efforts are not effective enough to overcome air pollution and also the lack of public awareness of the dangers of air pollution.

Air quality monitoring is currently needed to evaluate the degree of air pollution objectively and predict pollutant concentrations accurately [13]. Currently, air quality monitoring to measure particulate concentrations have been done by [14] [15]. However, these methods tend to be too expensive and too sophisticated for most people to use. The use of sensor nodes in air quality monitoring in research [16] have been done to measure air quality parameters. The proposed mobile sensor network can handle dynamic coverage successfully, but it is expensive to develop. The use of low cost sensors to measure air pollution have been done in research [17] [18]. However, a survey that done by [19] stated that the accuracy of using low-cost sensor is more prone to error, so it is less suitable if it is often used in harsh environments. Therefore, to improve existing air quality monitoring efforts, an accurate and efficient modeling method or technique is needed which can classify air quality.

Several research for classification cases have been done and usually use an algorithms approach machine learning. Machine Learning is one of computer science that is capable of performing a task without being explicitly programmed [3]. There are several machine learning algorithms that can be used to classify air quality, such as Random Forest, Decision Tree, Support Vector Machine, and Deep Belief Network [20]. In the research [21] used the Artificial Neural Network method to classify the Air Pollution Index in Shanghai, but to improve the better results the ANN method needs to be developed. The Decision Tree and Naïve Bayes methods in this research [22][23] were used to improve the accuracy in classifying air quality, but the classification accuracy results were obtained only 85.71% and 86,663%. Recently, the PCA-Neural Network method was used to forecast the Air Pollution Index in Delhi, but the technique of approaching this method often not the best for air quality forecast for practical and operational reasons [24].

The use of an accurate model in classifying air pollution is very influential in decision making. This is useful in developing strategies for the environment, as well as protection in reducing damage to ecosystems and protect human health [25]. Several research in [21][22][23][24] have not been effective enough in classifying air quality in terms of time, technique, and level of accuracy. Therefore, a more precise and better performance method is needed in classifying air quality. The Support Vector Machine method is proposed to classify air quality, because it has excellent performance in terms of accuracy, complexity, and can solve small, nonlinear and high dimensional estimation problems [26][27]. This is evidenced by research [28], [29], [30], [31], and [32] where the classification performance with high accuracy performance is above 90%, 93.2%, 97.01%, 98, 53%, and 100%.

Support Vector Machine are not only known for its good accuracy performance, but the feasibility of applying SVM is also examined as evidence of the superiority of this method in various classification cases. Research on [33] has successfully applied the SVM method to predict solar radiation in various climatic regions of China. The comparison of method for air quality forecasting was done in [34], however from the performance side with different time series it was stated that SVM was superior in predicting air quality parameters. Furthermore, implementing SVM provides a flexible and scalable tool for implementing sophisticated solutions with dynamic and non-linear data in terms of methods, techniques and alternatives [35]. Another advantages of SVM is also stated in various cases in research [36] and [37].

In this research, the feasibility of implementing SVM and the advantages offered by the SVM method will be used to classify air quality from the monitoring results based on five air pollutant parameters. The classification results with Support Vector Machine able to describe the current air quality conditions. The goal of this research

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

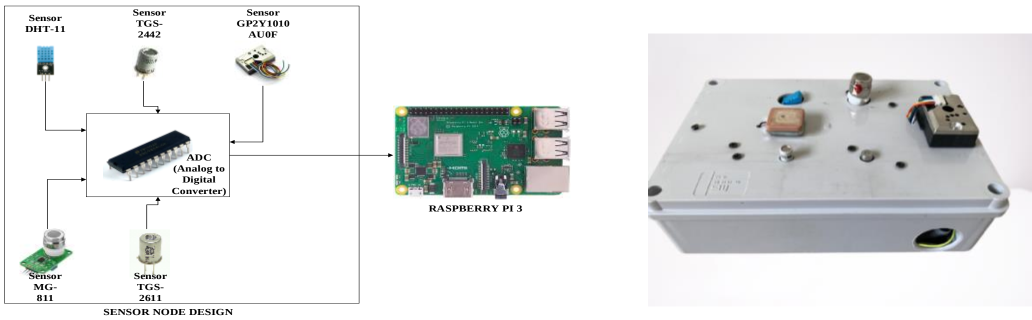
is to apply the Support Vector Machine method in classifying air quality and see the classification accuracy performance using Support Vector Machine method.

2. AIR QUALITY CLASSIFICATION

To get good air quality classification performance, sensor technology and machine learning with appropriate classification method are needed. Several research on air quality monitoring was done by [3], [38], [39], but this method can only be used in non-linear or linear problems. SVM is a method that can be used in classification problems in both linear and non-linear cases [40]. Comparisons between classification methods have been done in various air quality monitoring research [40], [41], [42], and [43], but in terms of data computation and classification performance, SVM is superior. A brief explanation of them is given as follows:

2.1 AIR QUALITY MONITORING SYSTEM DESIGN

This section will describe the results of the research which provides a comprehensive research of the performance of the air quality monitoring system. The block diagram proposed in the monitoring system consists of a sensing unit, the controller unit, and the output data from the sensor shown in Figure 1(a). The development of sensor technology has led to the rapid growth in being able to detect, measure and collect information from the real world (weather conditions, quality air) [44]. Therefore, the use of integrated gas sensors is an alternative way of monitoring and characterize air quality [45].



(a) (b)

FIGURE 1. (a) Block Diagram of Air Quality Monitoring System Design, (b) Hardware

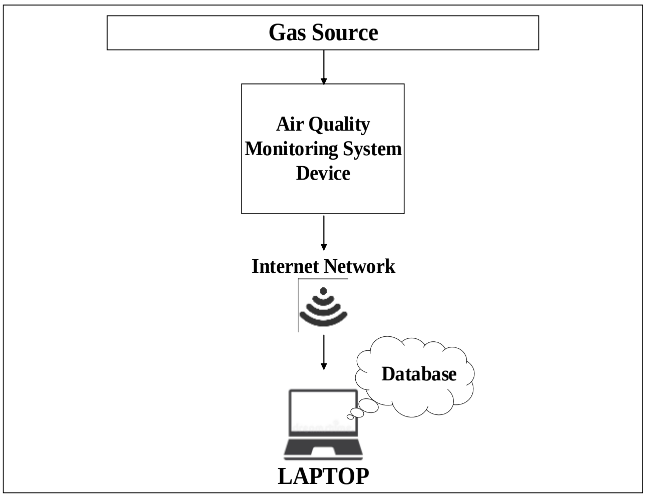


FIGURE 2. Air Quality Monitoring System to get the data

In this research, using an "Air Quality Monitoring System" which consists of five sensors, namely TGS-2442, TGS-2611, MG-811, GP2Y1010AU0F, and DHT-11. The five sensors will be connected to the Analog Digital Converter (ADC) which will enter the Raspberry Pi input for processing (See Fig. 1a). The sensors will detect pollutants around the system, so that the sensor readings are generated. Each sensor reading that is read by the system will be classified using SVM and the classified data will be sent to the database using the internet network, so that data information that has been sent to the database can be viewed from laptop (See Fig. 2).

2.2 SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a machine learning algorithm that can help solve big data classification problems [46]. It was first introduced by Vapnik in 1979 [47]. To achieve this goal, Figure 3 is divided into three parts, namely training data, machine learning algorithms and models. In this research, this machine learning uses the SVM algorithm based on training data to form a classification model. The classification model is used to classify the tested sample, namely the output data from the sensor. Machine learning for air quality prediction is an efficient and convenient way to solve several related environmental problems [45], and automatically learnt the information and easily recognizes the complex patterns [48].

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

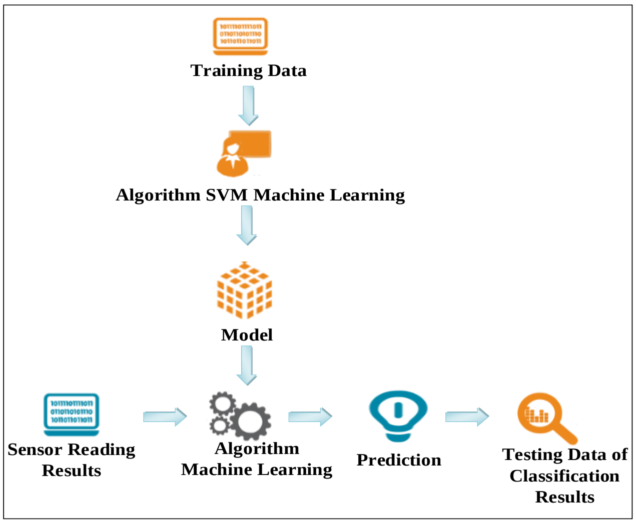


FIGURE 3. Block Diagram of the SVM Classifier

Classification problems in SVM can be limited to consideration of two class problems without loss of generality [49]. The goal is to separate the two classes, so as to produce a classifier that will work well. Figure 4 shows that there are many possible linear classifiers that can separate the data [2]. This linear classification is called the optimal separator hyperplane. However, the goal of SVM is to draw an optimal line to separate the two classes with maximum separation margin. This idea can be conceptualized to create an optimization problem [47].

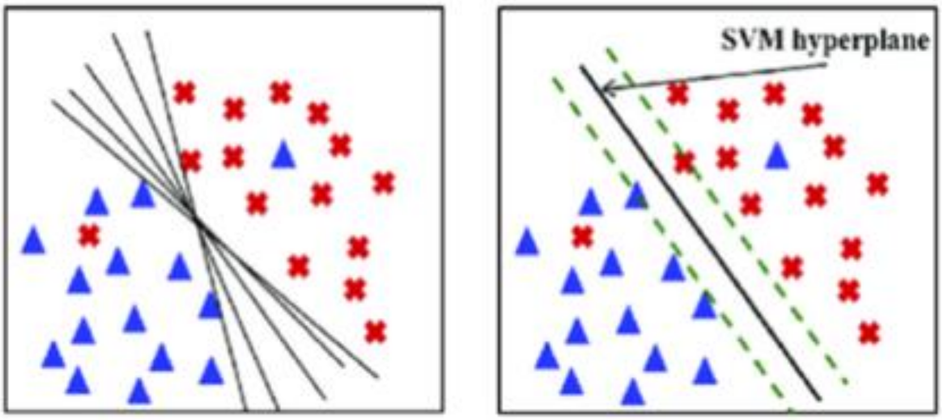


FIGURE 4. Optimal Separating Hyperplane

Based on the training data set in Figure 4 with input ∈ and output ∈ ±1:

$$(x_i, y_i) \in R^1 * \{\pm 1\}, i = 1, \ldots\ldots N \quad (1)$$

There are two classes, where '+1' represents one class, and '-1' represents another class. After training, it is expected that the decision functions are given by:

$$f_{w,b}(x) = \mathrm{sgn}(w.x + b) \quad (2)$$

Here w is the coefficient vector and b is the bias of the hyperplane. 'sgn' represents the bipolar sign function. Ideally, the following conditions should be satisfied by the hyperplane of the classifier.

$$y_i[w.x_i + b] \geq 1,\ i = 1,2, \ldots\ldots.., N \qquad (3)$$

Among all satisfying separating hyperplane, the hyperplane with the maximum distance to the nearest point is considered as the optimal separating hyperplane. Depends on:

$$y_i[w.x_i + b] \geq 1 - e_i,\ i = 1,2, \ldots\ldots .., N \qquad (4)$$

where e_i is the variable slack> 0 which allows to manage when the ideal hyperplane (3) is not possible.

The feasibility of implementing SVM is examined through a comparison of the performance of three kernels namely linear, polynomial, and RBF [31]. The SVM algorithm does not consider whether the data is balanced or not. However, the problem of unbalanced data is a challenging problem to build a strong classification models [50]. Kernel parameters have a great influence on the complexity and performance of the prediction model. Hence, model selection in SVM involves kernel parameter [50]. Therefore, machine learning algorithms in classification are very important in order for decision making to get good predictive results (See Fig. 2). In this research, the air quality monitoring system will detect five pollutant pollutant parameters namely CO, CO_2, HC, dust/PM_{10}, and temperature through sensors contained in the "Air Quality Monitoring System". The five pollutants are used because they are most contribute for air pollution problems. The use of the SVM algorithm here aims to classify each sensor reading into three classifications, namely Normal, Moderate and Hazardous and to see the classification performance obtained by SVM. Overall, the results of this research are only presented in the form of simulations and classification in the real experiment.

2.3 PREPROCESSING DATA

At the data collection stage, the observed data were five air pollutants consisting of CO, CO_2, HC, dust/PM_{10}, and temperature. This data is obtained during testing from the results of air quality monitoring in the form of sensor readings. After the sensor readings are obtained, an image of the input space of the five pollutants will be generated which will then be processed using the Support Vector Machine algorithm.

TABLE 1
Sensor Readings Data Structure

Data testing to-	Air Quality Parameters				
	CO (ppm)	CO_2 (ppm)	HC (ppm)	PM_{10} (µg / m3)	Temperature (° C)
1.	A1,1	B1,1	C1,1	D1,1	E1,1
2.	A1,2	B1,2	C1,2	D1,2	E1,2
…	…	…	…	…	…
100	…	…	…	…	…

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

Based on the results of the sensor readings obtained in Table 1, then the data processing was did using the model that has been generated (See Fig. 3) as a reference in classifying the testing data. The Support Vector Machine algorithm will recognize patterns in the testing data to find the best hyperplane in the input space. After obtaining the hyperplane, it can be seen the classification results of the separation of two different classes.

2.4 CLASSIFICATION PERFORMANCE

In this research, the results of the classification can be evaluated by counting the number of correct predictions and the number of wrong predictions from the test results. This evaluation aims to see the accuracy of the classification of the Support Vector Machine method in classifying air quality. Classification accuracy results in a presentation of the accuracy obtained by statistical calculations. Classification accuracy indicates the effectiveness of the classifier as a whole. The higher the accuracy value, the better the classifier's performance in classifying the data.

3. RESULTS AND DISCUSSION

3.1 TRAINING DATA

The training data that was used is a datasheet that contains the range of the reading values for each sensor. Each sensor reading value will be labeled into three classification parameters, namely Normal, Moderate, and Hazardous. After the training data is labeled, a learning process was did using the SVM algorithm to produce a model.This model will be used repeatedly to classify air quality during testing.

To ensure the efficiency of the model formed, a model suitability test was did using kernel functions. It aims to train the model and classify the training data prior to classification on the testing data, so that optimal classification accuracy performance is obtained. Table 2 shows that the linear kernel function is better than other kernel functions, where the classification accuracy results with the linear kernel function on the training data are 1.0, which means 100%.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

TABLE 2
Performance Results of Kernel Function Metrics

Kernel	Class	Performance Metrics		
		Precision	*Recall*	*f1-score*
Linear	0.0	1.00	1.00	1.00
	1.0	1.00	1.00	1.00
	2.0	1.00	1.00	1.00
Polynomial	0.0	0.90	1.00	0.95
	1.0	1.00	0.88	0.94
	2.0	1.00	1.00	1.00
Gaussian	0.0	0.92	1.00	0.96
	1.0	1.00	0.88	0.93
	2.0	0.93	1.00	0.97
Sigmoid	0.0	0.53	0.40	0.46
	1.0	0.00	0.00	0.00
	2.0	0.20	0.56	0.30

3.2 CLASSIFICATION RESULT IN REAL EXPERIMENT

After obtaining the model from the training data process, then the testing process was did in realtime. During the monitoring process, all sensor readings read by the system will be sent and stored in a database via the internet network. Table 3 is a sample of some of the sensor readings obtained during the air quality monitoring process.

In Table 3 for the first testing data, the CO reading is obtained 65.05427165161836 ppm, the CO_2 reading is 389.78644953558114 ppm, the result of the HC reading is 477.6918101185048 ppm, the PM_{10} reading is 32.45577586533722 µg/m^3, and the temperature reading is 31°C. From the results of the first sensor data testing, the weather conditions are classified as Normal (0.0). Then in Table 3 for the fourth testing data, the CO reading is 54.64650086317481 ppm, the CO_2 reading is 1843.7651698689584 ppm, the result of the HC reading is 417.27261578493238 ppm, the PM10 reading is 28.393486747808696 µg/m^3, and the temperature reading is 31°C. From the results of sensor readings on the fourth testing data, the weather conditions are classified as Moderate (1.0). However, there is a significant difference in the results of CO_2 readings between the first testing data and the fourth testing data, where the fourth test data obtained a fairly high CO_2 reading, namely 1843.7651698689584 ppm. This is what causes the fourth testing data to be classified as Moderate (1.0) compared to the first testing data which results in a smaller CO_2 reading, which is equal to389.78644953558114 ppm. In addition, this increase in CO_2 occurs due to the large amount of pollution from vehicles passing by at that location, so that the system detects a high enough CO_2 level and causes it to be classified as Moderate (1.0).

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

TABLE 3
Sensor Reading Results

Data testing to-	Air Quality Parameters					SVM Classification
	CO (ppm)	CO_2 (ppm)	HC (ppm)	PM_{10} (µg / m3)	Temperature (ºC)	
1.	65.05427165161836	389.78644953558114	477.6918101185048	32.45577586533722	31	Normal (0.0)
2.	48.27732325043189	600.05560619438756	333.75041724491183	16.22788793266861	33	Moderate (1.0)
3.	48.18931516672642	607.92242444088123	334.156870071636	14.526405580619478	31	Moderate (1.0)
4.	54.64650086317481	1843.7651698689584	417.27261578493238	28.393486747808696	31	Moderate (1.0)
5.	52.0975259949802	243.77544906884802	312.964680575164	16.844675285286424	31	Normal (0.0)
6.	66.07125395221487	257.68703328128635	372.7642334668252	23.16142851726883	29	Moderate (1.0)
7.	43.95514847289677	428.44144180750584	330.6506593298363	23.629336164082343	32	Normal (0.0)
8.	49.68219303106359	317.32853440029663	303.6470019520548	22.33195587064488	32	Normal (0.0)
9.	48.5250497082695	678.7580469789775	335.4472391373159	14.143572051408425	30	Moderate (1.0)
10.	14.837511001010464	431.4359620328103	382.4009276896286	14.12230352200781	30	Normal (0.0)
11.	63.94276215000489	342.61239668414873	474.28549645179424	30.669219395685634	33	Normal (0.0)
12.	65.53994589132631	308.40598893441415	352.5395013478226	22.28941881184365	29	Normal (0.0)
13.	49.307343785651426	402.465338270746	359.83673710446675	22.119270576638737	32	Normal (0.0)
14.	56.33169268880994	405.2945259921079	336.60600299551584	23.310308223073132	31	Normal (0.0)
15.	55.891652270282606	377.0183129726944	294.3677863472507	22.650983811654093	33	Normal (0.0)

The increase in CO_2 continues to increase in line with the increasing number of vehicles passing by at that location. The amount of pollution from these vehicles is one factor in increasing CO_2. The 12th testing data shows the results of testing data classified as Normal (0.0) where the CO_2 and temperature readings obtained are relatively small, namely 308.40598893441415 ppm and 29°C. This is because the number of vehicles available is small and the sky is cloudy, causing a drop in temperature at that location.

In Table 4 is a sample of some of the sensor readings that have misclassification. From Table 4, classification errors occur because the temperature readings obtained are dominant outside the normal limits. This can also be proven through the weather conditions observed at https://weather.com. In addition, misclassification also occurred in the first testing data where the CO_2 reading results obtained were 6005.21308876887 ppm. At that time, the sensor readings obtained were classified as Moderate (1.0). Classification errors obtained during testing are caused due to system instability, so that the system is unable to produce the correct sensor reading values. This also causes the sensor readings to not be classified properly. Overall, the total of 2045 testing data obtained during testing and the classification using the Support Vector Machine method was proven have a good performance of accuracy, namely 95.02% with a misclassification of 4.98%.

TABLE 4
Misclassification Results

Data testing to-	Air Quality Parameters					SVM Classification
	CO (ppm)	CO_2 (ppm)	HC (ppm)	PM_{10} (µg / m3)	Temperature (ºC)	
1.	9.9057987 54848593	6005.2130 8876887	416.889462 24154716	20.949501 459604956	26	Moderate (1.0)
2.	65.993024 54447668	277.05383 51905071	369.751877 71522377	22.778594 98805778	13	Normal (0.0)
3.	35.871443 00661691	306.56379 266838303	296.946641 48624645	24.565151 457709366	17	Normal (0.0)
4.	45.483881 48244727	664.50400 884908044	308.865269 49902327	23.905827 04629033	32	Normal (0.0)
5.	37.465367 189282574	428.02316 56302729	409.115478 33585604	124.65041 233416996	35	Normal (0.0)

3.3 CLASSIFICATION IN SIMULATION

The following is a simulation of the test results which can be seen in Figures 5, 6 and 7. Figures 5, 6, 7 are the simulation results of the SVM classification from the testing data. Figure 5 describes normal data and unnormal data. So, basically SVM method can only distinguish two classes, namely +1 and -1. In Figure 5, SVM distinguishes which class is classified as normal (+1) and which is classified as unnormal (-1). If the data distribution is included in the normal area, it means that the sensor readings obtained in that area are classified as normal. However, if the data distribution is included in the unnormal area, then proceed to the next step, which is to check again whether the sensor readings in that area are classified as moderate or unmoderate.

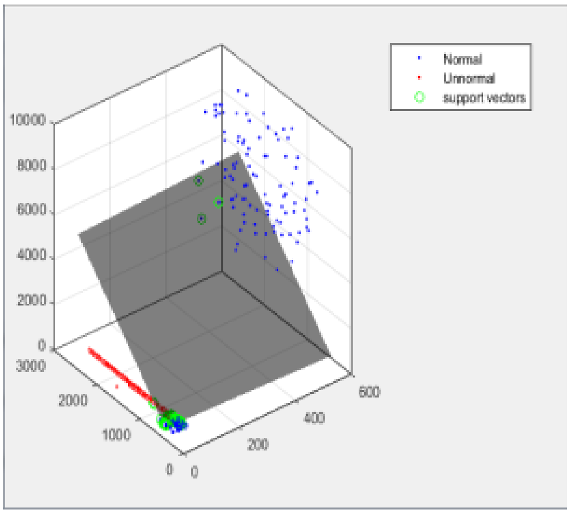


FIGURE 5. Normal classification simulation results on testing data

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

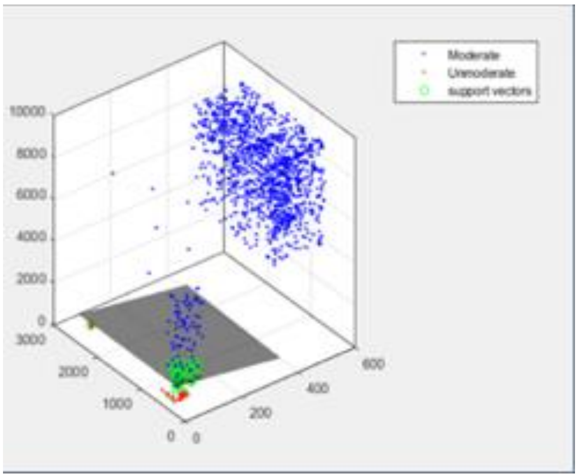


FIGURE 6. Simulation results of the Moderate classification on the testing data

Figure 6 shows the results of the data distribution for moderate and unmoderate use of the SVM method. It can be seen that the distribution of data is included in the moderate area, meaning that the sensor readings obtained in that area are classified as moderate. However, if the data distribution is included in an unmoderate area, then proceed to the next step, which is to check again whether the sensor readings in that area are classified as hazardous or unhazardous as shown in Figure 7.

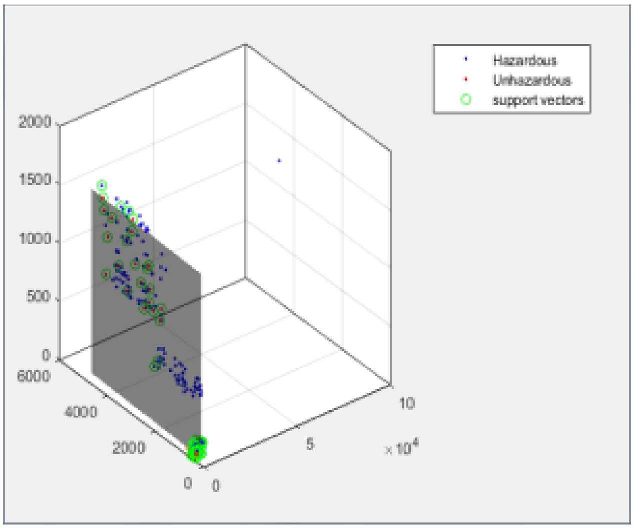


FIGURE 7. Simulation results of Hazardous classification on testing data

4. CONCLUSION

The results of this research can be concluded that the Support Vector Machine method has been successful in classifying air quality from the sensor readings. The stability of the sensor can affect the success in data classification. The classification accuracy performance obtained was 95.02%. This is evidenced by the results of the test data simulation which are able to separate two different data classes. In this research, the use of sensors that are not properly conditioned can cause the system to produce incorrect sensor readings, which can lead to misclassification in testing.

ACKNOWLEDGEMENTS

Authors thank to the Kementrian Pendidikan dan Kebudayaan (Kemendikbud) Indonesia, RistekBrin and Sriwijaya State Polytechnic of for their financial support in Grants Project. Our earnest gratitude also goes to all researchers in Telecommunication and Signal and Control Laboratory, Electrical Engineering, Polytechnic Sriwijaya who provided companionship and sharing of their knowledge.

REFERENCES

[1] I. Suryati, H. Khair, and D. Gusrianti, "Analysis of air quality index distribution of PM10 and O3 concentrations in ambient air of Medan City, Indonesia," *J. Phys. Sci.*, vol. 29, pp. 37–48, 2018, doi: 10.21315/jps2018.29.s3.5.

[2] N. L. Husni, A. S. Handayani, S. Nurmaini, and I. Yani, "Odor classification using Support Vector Machine," *ICECOS 2017 - Proceeding 2017 Int. Conf. Electr. Eng. Comput. Sci. Sustain. Cult. Herit. Towar. Smart Environ. Better Futur.*, no. December 2018, pp. 71–76, 2017, doi: 10.1109/ICECOS.2017.8167170.

[3] G. K. Kang, J. Z. Gao, S. Chiao, S. Lu, and G. Xie, "Air Quality Prediction: Big Data and Machine Learning Approaches," *Int. J. Environ. Sci. Dev.*, vol. 9, no. 1, pp. 8–16, 2018, doi: 10.18178/ijesd.2018.9.1.1066.

[4] A. S. Handayani, N. L. Husni, R. Permatasari, and C. R. Sitompul, "Implementation of Multi Sensor Network as Air Monitoring Using IoT Applications," *Int. Tech. Conf. Circuits/Systems, Comput. Commun. (ITC-CSCC), JeJu, Korea*, pp. 1–4, 2019, doi: 10.1109/ITC-CSCC.2019.8793407.

[5] B. Haryanto, "Climate change and urban air pollution health impacts in Indonesia," in *Climate Change and Air Pollution*, Springer, 2018, pp. 215–239.

[6] S. N. Kane, A. Mishra, and A. K. Dutta, "Preface: International Conference on Recent Trends in Physics (ICRTP 2016)," *J. Phys. Conf. Ser.*, vol. 755, no. 1, 2016, doi: 10.1088/1742-6596/755/1/011001.

[7] J. Lou, Y. Wu, P. Liu, S. H. Kota, and L. Huang, "Health effects of climate change through temperature and air pollution," *Curr. Pollut. Reports*, vol. 5, no. 3, pp. 144–158, 2019.

[8] Peraturan Pemerintah, "Peraturan Pemerintah no. 41 tentang Pengendalian Pencemaran udara," *Peratur. Pemerintah no. 41 tentang Pengendali. Pencemaran Udar.*, no. 1, pp. 1–5, 1999, doi: 10.1016/j.aquaculture.2007.03.021.

[9] S. Meerow and J. P. Newell, "Spatial planning for multifunctional green infrastructure: Growing resilience in Detroit," *Landsc. Urban Plan.*, vol. 159, pp. 62–75, 2017, doi: https://doi.org/10.1016/j.landurbplan.2016.10.005.

[10] D. Sofia, F. Gioiella, N. Lotrecchiano, and A. Giuliano, "Mitigation strategies for reducing air pollution," *Environ. Sci. Pollut. Res.*, vol. 27, no. 16, pp.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

19226–19235, 2020, doi: 10.1007/s11356-020-08647-x.

[11] M. Farda and C. Balijepalli, “Exploring the effectiveness of demand management policy in reducing traffic congestion and environmental pollution: Car-free day and odd-even plate measures for Bandung city in Indonesia,” *Case Stud. Transp. Policy*, vol. 6, no. 4, pp. 577–590, 2018, doi: https://doi.org/10.1016/j.cstp.2018.07.008.

[12] H. Igliński and M. Babiak, “Analysis of the Potential of Autonomous Vehicles in Reducing the Emissions of Greenhouse Gases in Road Transport,” *Procedia Eng.*, vol. 192, pp. 353–358, 2017, doi: 10.1016/j.proeng.2017.06.061.

[13] Z. Yang and J. Wang, “A new air quality monitoring and early warning system: Air quality assessment and air pollutant concentration prediction,” *Environ. Res.*, vol. 158, pp. 105–117, 2017, doi: https://doi.org/10.1016/j.envres.2017.06.002.

[14] E. Ruppecht, M. Meyer, and H. Patashnick, “The tapered element oscillating microbalance as a tool for measuring ambient particulate concentrations in real time,” *J. Aerosol Sci.*, vol. 23, pp. 635–638, 1992, doi: https://doi.org/10.1016/0021-8502(92)90492-E.

[15] H. Hauck *et al.*, “On the equivalence of gravimetric PM data with TEOM and beta-attenuation measurements,” *J. Aerosol Sci.*, vol. 35, no. 9, pp. 1135–1149, 2004, doi: https://doi.org/10.1016/j.jaerosci.2004.04.004.

[16] A. Marjovi, A. Arfire, and A. Martinoli, “High resolution air pollution maps in urban environments using mobile sensor networks,” *Proc. - IEEE Int. Conf. Distrib. Comput. Sens. Syst. DCOSS 2015*, pp. 11–20, 2015, doi: 10.1109/DCOSS.2015.32.

[17] H. Li, Y. Gao, P. Shi, and H.-K. Lam, “Observer-based fault detection for nonlinear systems with sensor fault and limited communication capacity,” *IEEE Trans. Automat. Contr.*, vol. 61, no. 9, pp. 2745–2751, 2015.

[18] L. Spinelle, M. Gerboles, M. G. Villani, M. Aleixandre, and F. Bonavitacola, “Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO2,” *Sensors Actuators B Chem.*, vol. 238, pp. 706–715, 2017, doi: https://doi.org/10.1016/j.snb.2016.07.036.

[19] B. Maag, Z. Zhou, and L. Thiele, “A Survey on Sensor Calibration in Air Pollution Monitoring Deployments,” *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4857–4870, 2018, doi: 10.1109/JIOT.2018.2853660.

[20] S. Y. Muhammad, M. Makhtar, A. Rozaimee, A. A. Aziz, and A. A. Jamal, “Classification model for water quality using machine learning techniques,” *Int. J. Softw. Eng. its Appl.*, vol. 9, no. 6, pp. 45–52, 2015, doi: 10.14257/ijseia.2015.9.6.05.

[21] D. Jiang, Y. Zhang, X. Hu, Y. Zeng, J. Tan, and D. Shao, “Progress in developing an ANN model for air pollution index forecast,” *Atmos. Environ.*, vol. 38, no. 40, pp. 7055–7064, 2004, doi: https://doi.org/10.1016/j.atmosenv.2003.10.066.

[22] B. Sugiarto and R. Sustika, “Data classification for air quality on wireless sensor network monitoring system using decision tree algorithm,” in *International Conference on Science and Technology-Computer (ICST)*, 2016, pp. 172–176, doi: 10.1109/ICSTC.2016.7877369.

[23] R. W. Gore and D. S. Deshpande, "An approach for classification of health risks based on air quality levels," in *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)*, 2017, pp. 58–61.

[24] A. Kumar and P. Goyal, "Forecasting of air quality index in Delhi using neural network based on principal component analysis," *Pure Appl. Geophys.*, vol. 170, no. 4, pp. 711–722, 2013.

[25] B. Wolterbeek, "Biomonitoring of trace element air pollution: principles, possibilities and perspectives," *Environ. Pollut.*, vol. 120, no. 1, pp. 11–21, 2002, doi: https://doi.org/10.1016/S0269-7491(02)00124-0.

[26] J. Li, B. Zhang, and J. Shi, "Combining a genetic algorithm and support vector machine to study the factors influencing CO2 emissions in Beijing with scenario analysis," *Energies*, vol. 10, no. 10, 2017, doi: 10.3390/en10101520.

[27] N. S. A. Yasmin, N. A. Wahab, A. N. Anuar, and M. Bob, "Performance comparison of SVM and ANN for aerobic granular sludge," *Bull. Electr. Eng. Informatics*, vol. 8, no. 4, pp. 1392–1401, 2019, doi: 10.11591/eei.v8i4.1605.

[28] Y. Kim and H. Ling, "Human Activity Classification Based on Micro-Doppler Signatures Using a Support Vector Machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–337, 2009, doi: 10.1109/TGRS.2009.2012849.

[29] L. C. Wu *et al.*, "Detection of American Football Head Impacts Using Biomechanical Features and Support Vector Machine Classification," *Sci. Rep.*, vol. 8, no. 1, pp. 1–14, 2018, doi: 10.1038/s41598-017-17864-3.

[30] S. W. Purnami, S. P. Rahayu, and A. Embong, "Feature selection and classification of breast cancer diagnosis based on support vector machines," in *International Symposium on Information Technology*, 2008, pp. 1–6, doi: 10.1109/ITSIM.2008.4631603.

[31] K. Polat and S. Güneş, "Breast cancer diagnosis using least square support vector machine," *Digit. Signal Process.*, vol. 17, no. 4, pp. 694–701, 2007, doi: https://doi.org/10.1016/j.dsp.2006.10.008.

[32] A. E. Minarno, F. D. S. Sumadi, H. Wibowo, and Y. Munarko, "Classification of batik patterns using K-nearest neighbor and support vector machine," *Bull. Electr. Eng. Informatics*, vol. 9, no. 3, pp. 1260–1267, 2020, doi: 10.11591/eei.v9i3.1971.

[33] J. Fan, X. Wang, F. Zhang, X. Ma, and L. Wu, "Predicting daily diffuse horizontal solar radiation in various climatic regions of China using support vector machine and tree-based soft computing models with local and extrinsic climatic data," *J. Clean. Prod.*, vol. 248, p. 119264, 2020, doi: https://doi.org/10.1016/j.jclepro.2019.119264.

[34] W. Z. Lu and W. J. Wang, "Potential assessment of the 'support vector machine' method in forecasting ambient air pollutant trends," *Chemosphere*, vol. 59, no. 5, pp. 693–701, 2005, doi: 10.1016/j.chemosphere.2004.10.032.

[35] A. Sotomayor-Olmedo, M. A. Aceves-Fernández, E. Gorrostieta-Hurtado, C. Pedraza-Ortega, J. M. Ramos-Arreguín, and J. E. Vargas-Soto, "Forecast Urban Air Pollution in Mexico City by Using Support Vector Machines: A Kernel Performance Approach," *Int. J. Intell. Sci.*, vol. 03, no. 03, pp. 126–135, 2013, doi: 10.4236/ijis.2013.33014.

[36] K. Aoyagi, H. Wang, H. Sudo, and A. Chiba, "Simple method to construct process maps for additive manufacturing using a support vector machine," *Addit. Manuf.*, vol. 27, pp. 353–362, 2019, doi:

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

https://doi.org/10.1016/j.addma.2019.03.013.
[37] J. Liu and E. Zio, "Integration of feature vector selection and support vector machine for classification of imbalanced data," *Appl. Soft Comput.*, vol. 75, pp. 702–711, 2019, doi: https://doi.org/10.1016/j.asoc.2018.11.045.
[38] I. N. Athanasiadis, V. G. Kaburlasos, P. A. Mitkas, and V. Petridis, "Applying Machine Learning Techniques on Air Quality Data for Real-Time Decision Support 1 Introduction 2 Decision support systems for assessing air quality in real – time 3 The σ – FLNMAP Classifier," *First Int. Symp. Inf. Technol. Environ. Eng.*, no. June, pp. 2–7, 2003.
[39] K. Y. Ng and N. Awang, "Multiple linear regression and regression with time series error models in forecasting PM 10 concentrations in Peninsular Malaysia," *Env. Monit Assess*, 2018, doi: 10.1007/s10661-017-6419-z.
[40] S. Bedoui, S. Gomri, H. Samet, and A. Kachouri, "A Prediction Distribution of Atmospheric Pollutants using Support Vector Machines, Discriminant Analysis and Mapping Tools (Case study: Tunisia)," *Polym. J.*, vol. 2, pp. 11–23, 2016, doi: 10.7508/PJ.2016.01.002.
[41] W. Wang, C. Men, and W. Lu, "Online prediction model based on support vector machine," *Neurocomputing*, vol. 71, no. 4, pp. 550–558, 2008, doi: https://doi.org/10.1016/j.neucom.2007.07.020.
[42] X. Li, J. Sha, and Z. liang Wang, "A comparative study of multiple linear regression, artificial neural network and support vector machine for the prediction of dissolved oxygen," *Hydrol. Res.*, vol. 48, no. 5, pp. 1214–1225, 2017, doi: 10.2166/nh.2016.149.
[43] X. Leng *et al.*, "Prediction of size-fractionated airborne particle-bound metals using MLR, BP-ANN and SVM analyses," *Chemosphere*, vol. 180, pp. 513–522, 2017, doi: https://doi.org/10.1016/j.chemosphere.2017.04.015.
[44] S. Kumar and A. Jasuja, "Air quality monitoring system based on IoT using Raspberry Pi," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, 2017, pp. 1341–1346.
[45] T. M. Amado, J. C., and D. Cruz, "Development of Machine Learning-based Predictive Models for Air Quality Monitoring and Characterization," *TENCON 2018 - 2018 IEEE Reg. 10 Conf.*, pp. 0668–0672, 2018, doi: 10.1109/TENCON.2018.8650518.
[46] S. Suthaharan, "Machine learning models and algorithms for big data classification," *Integr. Ser. Inf. Syst*, vol. 36, pp. 1–12, 2016.
[47] Vapnik V. N., *The Nature of Statistical Learning Theory(2ed)*. .
[48] M. Nawir, A. Amir, N. Yaakob, and O. B. Lynn, "Effective and efficient network anomaly detection system using machine learning algorithm," *Bull. Electr. Eng. Informatics*, vol. 8, no. 1, pp. 46–51, 2019, doi: 10.11591/eei.v8i1.1387.
[49] F. M. Vieira, C. D. O. Bizarria, C. L. Nascimento, and K. T. Fitzgibbon, "Health monitoring using support vector classification on an auxiliary power unit," *IEEE Aerosp. Conf. Proc.*, no. April, 2009, doi: 10.1109/AERO.2009.4839655.
[50] A. Tharwat, "Behavioral analysis of support vector machine classifier with Gaussian kernel and imbalanced data," pp. 1–23, 2020, [Online]. Available: http://arxiv.org/abs/2007.05042.