Jurnal Teknologi

Full Paper

GAS SOURCE LOCALIZATION THROUGH DEEP LEARNING METHOD BASED ON GAS DISTRIBUTION MAP DATABASE

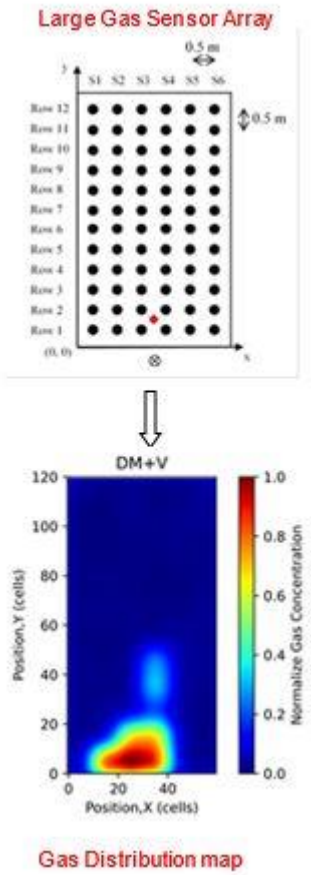
Zaffry Hadi Mohd Juffry[a], Kamarulzaman Kamarudin[a]*, Abdul Hamid Adom[a], Muhammad Fahmi Miskon[b], Ahmad Shakaff Ali Yeon[a], Abdulnasser Nabil Abdullah[a]

[a]Faculty of Electrical Engineering & Technology, Universiti Malaysia Perlis (UniMAP), 02600 Arau Perlis Malaysia
[b]Faculty of Electrical Engineering Universiti Teknikal Malaysia Melaka (UTeM), 76100 Durian Tunggal, Melaka Malaysia

Article history
Received
7 April 2023
Received in revised form
26 November 2023
Accepted
26 November 2023
Published Online
18 February 2024

*Corresponding author
kamarulzaman@unimap.edu.my

Graphical abstract



Abstract

The incident of harmful gas leakage can cause severe damage to the environment and several casualties to human beings while the gas localization system plays a major role in mitigating those causalities. With the advances in artificial intelligence technology, deep learning is able to enhance the accuracy of the gas localization system to locate the gas source. This paper proposes a gas localization system that utilizes three different deep learning models namely DNN, 1DCNN, and 2DCNN to locate the gas source within the gas map. The proposed method involves generating the gas distribution map through the large gas sensor array platform in real-world indoor scenarios. Those models are then trained using the collected database which allows for accurate prediction of the gas source location. The performance of each proposed deep learning model was compared to find the best model demonstrating the highest effectiveness in identifying gas leaks. The study has shown that the 1DCNN has the highest effectiveness in predicting the gas source in the range between 0.0 m to 0.3 m with 90.3% compared to the DNN and 2DCNN models.

Keywords: Gas source localization, gas distribution map, deep learning, harmful gas dispersion, mobile robot olfaction

Abstrak

Pengumpulan Insiden kebocoran gas berbahaya boleh menyebabkan kerosakan teruk kepada alam sekitar dan kecederaan kepada manusia. Sistem pencarian gas memainkan peranan utama untuk mengurangkan sebab-sebab tersebut. Dengan kemajuan dalam teknologi kecerdasan buatan pembelajaran mendalam mampu meningkatkan ketepatan sistem pencarian gas untuk mencari sumber gas. Kertas kerja ini mencadangkan sistem pencarian gas yang menggunakan tiga model pembelajaran mendalam yang berbeza iaitu dikenali sebagai DNN, 1DCNN dan 2DCNN untuk mencari sumber gas yang terdapat di dalam peta gas. Pendekatan yang dicadangkan melibatkan penjanaan peta penyebaran gas melalui platform susunan penderia gas yang besar dalam senario dunia sebenar. Model tersebut kemudian dilatih menggunakan pangkalan data yang dikumpul yang membolehkan ramalan tepat lokasi sumber gas. Prestasi setiap model pembelajaran mendalam yang dicadangkan telah dibandingkan untuk mencari model terbaik yang akan menunjukkan keberkesanan tertinggi dalam mengenal pasti kebocoran gas. Kajian ini telah menunjukkan bahawa 1DCNN mempunyai keberkesanan tertinggi dalam meramalkan sumber gas dalam julat antara 0.0 m hingga 0.3 m dengan 90.3% berbanding model DNN dan 2DCNN.

Kata kunci: Pencarian punca gas, peta penyebaran gas, pembelajaran dalam, penyebaran gas berbahaya, robot penghidu bergerak

1.0 INTRODUCTION

Gas source localization (GSL) is a critical research topic that aims to identify the origin of gas emissions, particularly in environments where hazardous gases may be released[1]. GSL can pose a significant challenge due to the slow process of molecular diffusion, where the released gas primarily spreads through airflow, creating a constantly fluctuating and complicated gas distribution shape, particularly in turbulent flows[2]. In addition, with the development of industries, the usage of dangerous chemical gaseous is increasing frequently. The release of dangerous chemicals in gas form is a common occurrence during storage and transportation processes[3] and can result in catastrophic events due to their chemical and physical properties. The toxic gases released during such events can form clouds that pose a threat to both indoor and outdoor environments[4]. Even though most industries are installed with the stationary gas sensor system yet the systems are unable to perform GSL tasks[5] since the system has low resolution and limited coverage which cannot capture the full extent of the gas distribution in a complex environment. Therefore, it is important to apply the GSL to the system in order to locate the gas source to prevent further harm.

In order for GSL to achieve success, several factors need to be considered. Firstly, there should be a gas sensor capable of detecting the target gas in the air. Secondly, an appropriate strategy should be employed to decide where the measurements need to be taken, and effective algorithms are necessary for processing the gas sensor signal and estimating the source location[6]. Currently, there are two different methods to collect the gas information to perform the GSL which by utilizing a mobile robot or installing stationary gas sensors.

In the mobile robot method, a mobile robot will be installed with gas sensors and carry out the sensors to the contaminated area and send all the information to the base station. This method allows the operator to stay in a secure location while searching for the harmful emission source. This method has several advantages such as low cost, ease of deployment, and flexibility. However, it is limited to mass data collection and is also time-consuming [7].

In the stationary gas sensors method, usually it needs a large number of sensors to develop the system. It consists of a network of gas sensors strategically placed in stationary positions to continuously monitor the air for the presence of certain harmful gases. Normally, this system is installed inside a particular building area that has a high potential for gas leakage to happen [8]. The system will be triggered when any of the sensors detect abnormal levels of gas concentration. In addition, the process of system development could potentially demand a significant amount of time and lack flexibility. However, they offer more comprehensive insight into the surroundings and can remain in place for consistent data collection [9].

In terms of the effective algorithm for processing the gas sensor signal, there is an effective method to represent the gas dispersion from the sensor reading which is known as the probabilistic method. This method treats the gas measurement as a probabilistic distribution and can be divided into two distinct groups which are infotaxis and Gas Distribution Map(GDM). However, this study will focus more on the GDM method.

The GDM is a process of creating a representation of how the gas disperses in the environment from a set of temporally and spatially distributed measurements of relevant gas information[10]. One of the GDM methods known as Kernel DM was proposed by A. Lilienthal & Duckett (2004) which implements a statistical method on the grid cells to create a gas concentration map[11]. It represented an estimation of the gas distribution mean and applied an extrapolation algorithm by convolving a single point of gas sensor reading to a nearby location (i.e. neighbourhood cells) by using a symmetrical Gaussian kernel. Then, further research was done to improve the method which is known as the Kernel DM+V algorithm that allows a larger dataset to be handled by a simpler learning procedure[12]. The grid cells that hold the highest variance value are assumed to be the gas source location. This work showed that the GDM method was able to declare the gas source location accurately.

In recent years, the learning method also has attracted lots of attention from researchers in the GSL field. This method allows the machine or system to perform and improve the GSL task based on the training process. There are several studies that focus on the utilization of artificial intelligence for gas localization assessment which uses machine learning such as Support vector machines (SVMs)[13] and kernel ridge regression[14]. Recently, H.Kim *et al.*, (2019) have explored the use of Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) and Feedforward Neural Network (FNN) models for identifying the source location of gas leaks[15]. By using Computational Fluid Dynamics (CFD) simulations, a total of 460 scenarios were generated and subsequently fed into both neural network models for training. Results indicate that LSTM-RNN demonstrated approximately 20% higher accuracy for leak spot location assessment compared to the FNN model.

Furthermore, C. Bilgera *et al.*, (2018) performed a combination of Convolutional Neural Networks (CNN) and LSTM models to estimate gas location in outdoor environments by utilizing gas sensor arrays to generate sequential datasets for model training[1]. The results indicate that Artificial Neural Networks (ANN) hold significant promise for gas localization tasks. Another study done by H. L. Yu *et al.*, (2022) attempted to predict the location of plastic burning using an ANN model, which included 16 inputs, 4 hidden, and 12 output neurons[16]. The study generated data such as burning location, wind speeds, and wind direction using CFD simulations, and then trained the ANN model. The findings indicate that the trained model achieved up to 85.71% validity with an average error of 3.86%.

There is also the application of DNN used to predict the location of the real gas source using a gas sensor

array in the indoor environment done by A. S. A. Yeon *et al.*, (2023). However, this study only focuses on the DNN model without comparing it with other deep learning and machine learning models. The gas sensor reading was directly fed into the DNN model for the gas source prediction. The study showed that the DNN model is able to locate the gas source with promising accuracy[17].

Based on the literature, numerous scholars have utilized synthetic gas map datasets generated through the computational fluid dynamic method while other studies focus on the real scenario of gas dispersion in order to train the machine learning model.

In this study, an array of stationary gas sensors located on a particular platform will be used to collect the real scenario of gas dispersion in the indoor environment. Then, the collected gas sensor information will be converted into the GDM by using the Kernel DM+V algorithm to generate a substantial dataset to train different deep learning models. Finally, this study will undertake an evaluation of the performance of each model in order to determine the most appropriate model for the GSL task.

2.0 METHODOLOGY

2.1 Data Collection Using Large Gas Sensor Array (LGSA) Platform

Notably, deep learning needs a large number of data to perform the learning process so that the model can be robust and accurate to predict the output[18]. In order to gather a significant amount of GDM data, it is necessary to employ an appropriate platform for conducting this research. For this work, an integrated mobile gas sensing testbed[19] located in the Centre of Excellence for Advance Sensor Technology (CEASTech), Universiti Malaysia Perlis was utilized as shown in Figure 1. It has several components where (A) ceiling-mounted cameras, (B) a gas sensor, (C) the communication module, and (D) the gas sensor board.

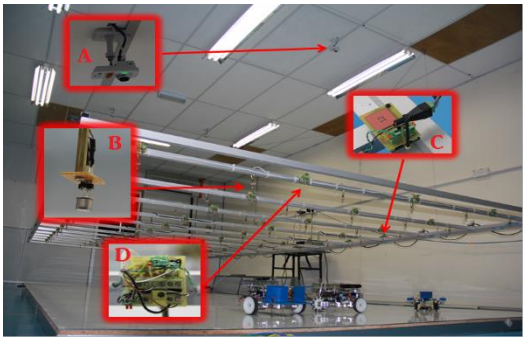


Figure 1 The integrated mobile robot testbed in CEASTech, UniMAP[19]

The designed platform is also known as a Large Gas Sensor Array (LGSA) since it has an array of gas sensors to capture the reading of released gas as shown in Figure 2. The red diamond shape illustrated the release point of the gas source. The platform has been installed with a total of 72 Figaro TGS2600 MOX gas sensors to cover a 3 m x 6 m testbed area. The TGS2600 was chosen since it is very sensitive to the ethanol vapor that is being used for this study. The 12-bit Analog-to-Digital Converter (ADC) (i.e., ADS7828) is paired with the gas sensor for the measurement process. The distance between each of the sensors is evenly separated by 0.5 m resulting in 12 rows and 6 columns respectively. The data was transmitted at a sample rate of 1 Hz to the PC-base station using a MEMSIC XM2100CA Wireless Sensor Network (WSN) transmitter (i.e., Wi-Fi module) to reduce the effect of voltage drop from a long wire connection.

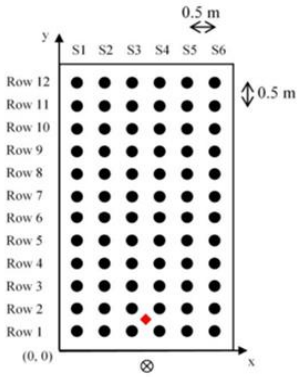


Figure 2 Gas sensor configuration in the large gas sensor array

All experiments were done using ethanol vapor which is produced by using a bubbler concept. The flow of clean air is forced into the bubbler by using an air pump so that the output of the bubbler outlet will emit ethanol vapor. Thus, the outlet of the bubbler acts as the gas source release point. There is a total of 60 different release points of gas source on the testbed platform. At each point, the gas was released at a flow rate of 0.05 kg/s for 30 minutes which follow the air pump flow rate specification. The LGSA system recorded all the readings from the gas sensors for 30 minutes and saved all the readings in the Excel file.

2.3 Data Pre-Processing

Before advancing to the next step, the collected readings from the gas sensor array should undergo the pre-processing data procedure to guarantee that all the data has been standardized and can be accommodated within the input layer of the DNN. The resistance of the gas sensor will vary as the gas concentration changes. As a result, the gas sensor's response can accurately identify the pattern of gas dispersion in the experimental area.

The response of the gas sensor, R_s able to calculate by using Equation 1. From this equation, the reading of sensor output voltage, V_{out} will be converted in terms of sensor response. V_{in} is the reference voltage for the gas sensor which is set to be 3.3 V while R_L is the load resistance for the sensor that has a value of 4.7 kΩ. The value of the load resistance is used at the optimum value which has been preferred by the manufacturer. This will allow the sensor to detect the gas even the gas at a low concentration. Then, the reading of the gas response is normalized (i.e., range value of [0,1]) by using Equation 2 by taking the aspect ratio of the gas response, and the baseline resistance reading.

$$R_s = \frac{V_{in}R_L}{V_{out}} - R_L \quad (1)$$

$$R_s = 1 - \frac{R_s}{R_0} \quad (2)$$

Next, the normalized reading of the sensor response has been converted into GDM by using the normal Kernel DM+V method with a kernel size of 0.5 m since it has been suggested that the estimations of gas source locations are often accurate in the range of 0.5 m to 1.5 m [20]. Another researcher also suggested that a reduced kernel size enhances the performance of the Kernel DM+V method [21].

The size of generated GDM is the same as the size of the testbed platform which is 3 m x 6 m while the dimension of each grid map is 5 cm x 5 cm, resulting in a total of 120 number of column and 60 rows. Therefore, there are 7200 total number of grid cells in the GDM, and each of the grid cells contains the reading of the normalized gas concentration. Figure 3 shows one of the GDM samples generated by the LGSA platform with the location of the gas release point at x = 30 y = 5 in terms of grid cells. From the figure, the red color indicates the high concentration of the gas while blue indicates the low concentration of gas reading.

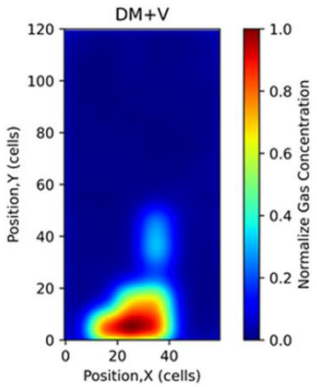


Figure 3 Gas distribution map (GDM)

There is a total of 5900 GDM samples produced from the LGSA platform and the samples were split into training and testing by the ratio of 0.33 to let the model have adequate testing data for evaluating the model performance and to prevent overfitting. Hence, the total amount of data for training and testing is 3953 and 1947 respectively. Since the model needs to be validated, the data for the validation purpose was taken from the testing dataset by a splitting ratio of 0.33. This will produce 1304 number of testing datasets and 643 number of validation datasets. All the dataset splitting process was done by the Scikit-learn package.

2.4 Deep Neural Network (DNN) Model

DNN is a type of machine learning that mimics the way the brain learns. It managed to learn from the input by using the probability method[22]. The reason DNN is chosen is because it increases the accuracy of the machine learning model. It also utilizes the hidden layer as a place to store and evaluate how significant one of the inputs to the output. It means all the information regarding the input's importance is stored in the hidden layer in terms of weight. The DNN used to supervise machine learning which means it's used labeled datasets to train the model and predict the outcome accurately.

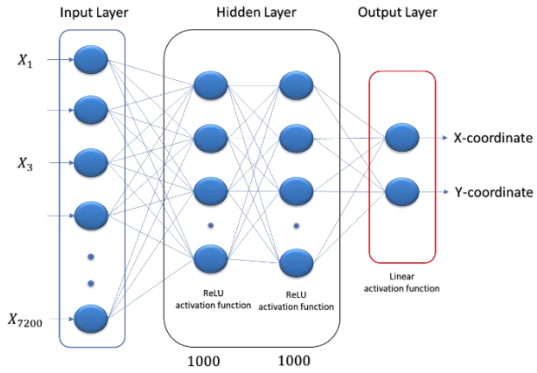


Figure 4 Deep neural network architecture

Figure 4 shows the complete model architecture of DNN that was developed by using Python 3.6 on the Google Collab platform. The DNN model consists of three major layers which are the input layer, hidden layer, and output layer. The GDM sample will be used as the input. Firstly, the two-dimensional (2D) (i.e., 60 x 120) array of the gas map was flattened into a one-dimensional (1D) array (i.e., 7200 x 1) shape before they fed through the input layer of DNN. The flattening was done to reduce the complexity of the model to be trained. This model consists of 7200 number of nodes in the input layer that represent the number of each grid cell in the flattened GDM sample. By entering the input layer with the reading of gas concentration from the flattened array, it will be performing the multiplying between the first layer node, x_i and synaptic weight, ω_i connected to the first hidden layer to get the result for the first hidden layer node as shown in Equation 3. The overall input for the

neuron is calculated through the weighted sum of the output signals that are received from the neurons in the previous layer:

$$u = \sum_{i-n}^{n} x_i \omega_i \tag{3}$$

Where x_i is shown in Equation 4:

$$x_i = X_n \tag{4}$$

Furthermore, the DNN model comprises two hidden layers utilizing the Rectified Linear Unit (ReLU) activation function. ReLU has become the default choice due to its ease of training and superior performance in various neural networks. The ReLU activation function defined all negative values of u to be 0 or otherwise to be a positive value of u itself as shown in Equation 5:

$$f(u) = \begin{cases} u \; for \; u \geq 0 \\ 0 \; for \; u \leq 0 \end{cases} \tag{5}$$

Based on the equation, u is the input ,and $f(u)$ is the output. As a result, if u is greater than 0, the derivative value is 1, and even if input data go through both hidden layers, the characteristics of the data remain similar without disappearing to the output layer.

Each hidden layer contains 1000 number of nodes. To prevent overfitting and long computational time, dropout function 0.2 was employed. The last component in the model which is the output layer consists of two nodes. The first node will represent the output of the gas source location on the x-axis while the second node will represent the output of the gas source location on the y-axis. Both of the output nodes used the linear activation function and the output will represent the location estimation of the gas source location in the continuous value in the range of 0 until 60 for the x-axis and in the range of 0 until 120 for the y-axis. Finally, during the training process, Mean Absolute Error (MAE) was implemented into the model as the loss and metric while the Adaptive Moment estimation (Adam) optimizer was chosen for the model optimizer to help the model optimize the learning process. The summary of the model is presented in Table 1.

Table 1 Summary of the DNN model

Layer	Output Shape	Number of Parameters
input	(120, 60)	0
flatten	(7200)	0
Dense 1	(1000)	7201000
Dense 2	(1000)	1001000
Dense X	1	1001
Dense Y	1	1001

2.5 One-dimensional (1D) Convolutional Neural Network (CNN)

In this study, 1DCNN has been trained using the GDM dataset for the gas source location prediction. It was chosen since the 1DCNN is able to identify the pattern in the 1D array and predict the output in terms of continuous value.

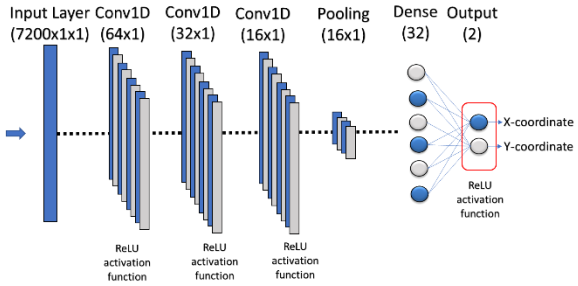


Figure 5 1DCNN model architecture

Figure 5 shows the architecture of the 1DCNN model to predict the location of the gas source. It consists of an input layer, 3 convolutional networks, a pooling layer, and 2 dense layers. For the input layer, it used the same method to insert the GDM data into the DNN model which means the 2D array (i.e., 60 x 120) of GDM is flattened to be a 1D array (i.e., 7200 x 1). Then, the data will pass through the first CNN layer which consists of 64 number of filters with a kernel size of 10 along with ReLU as the activation function. This layer takes the first layer as the input and transforms the sample into a 7191 x 64 array shape vector. After passing through the first CNN layer the sample passes through the dropout layer with a 0.1 rate that turns off 10% of the neurons to prevent overfitting. Then, it is followed by the second CNN layer consisting of 32 number of filters with a kernel size of 5 and utilizes ReLU as the activation function. The last CNN layer was consisting 16 number of layers with a kernel size of 3 and ReLU as the activation function. The last CNN layer was followed by max-pooling which has a pool size of 16. The utilization of max pooling helps to reduce the complexity and computational cost for the model to process the sample. Afterward, 2 dense layers were employed using the ReLU activation function to represent the location estimation of the gas source location in the continuous value in the range of 0 until 60 for the x-axis and in the range of 0 until 120 for the y-axis.

Finally, for the model compilation during the training process, Mean Absolute Error (MAE) was implemented into the model as the loss and metric while the Adaptive Moment estimation (Adam) optimizer was chosen for the model optimizer to help the model optimize the learning process. The summary of the model is presented in Table 2.

Table 2 Summary of the 1DCNN model

Layer	Output Shape	Number of Parameters
Conv1D	(7191, 64)	704
Dropout	(7191, 64)	0
Conv1D	(7187, 32)	10272
Conv1D	(7185, 16)	1552
MaxPooling1D	(2395, 16)	0
Flatten	38320	0
Dense	32	1226272
Dense	2	66

2.6 Two-dimensional (2D) Convolutional Neural Network (CNN)

The third deep learning model which is the 2DCNN model consists of three main layers which are a convolutional layer, a pooling layer, and a fully connected layer. The difference between the 2DCNN and 1DCNN is only the number of dimensions of each layer. The 2DCNN primarily emphasizes the analysis of data in 2D shapes, whereas the 1DCNN places greater emphasis on understanding the data's 1D shape.

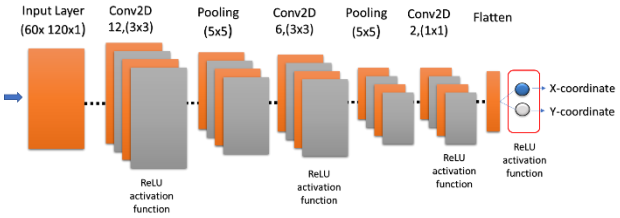


Figure 6 2DCNN model architecture

Figure 6 shows the model's architecture of 2DCNN for gas localization prediction. The model consists of 3 convolutional layers, 2 pooling layers 1 dense layer. For the input layer, the GDM dataset will not go through the flattening process as done in 1DCNN. The 60 x 120 array shape of the GDM was fed directly into the 2DCNN. Then, it passes through the first CNN layer which consists of 12 filters with a kernel size of 3 x 3 along with ReLU as the activation function. Next, max pooling reduces the number of dataset parameters by a size of 5 x 5. It is followed by the second convolutional layer which consists of 6 filters with a kernel size of 3 x 3 along with ReLU as the activation function. Then, the second pooling layer reduces the size of the dataset by a size of 5 x 5. Next, for the last convolutional layer, it utilizes two filters with a kernel size of 1 x 1 along with ReLU as the activation function. Afterward, after flattening the output from the convolutional layer, 2 dense layers were employed using the ReLU activation function to represent the location estimation of the gas source location in the continuous value in the range of 0 to 60 for the x-axis and in the range of 0 until 120 for the y-axis.

Finally, Mean Absolute Error (MAE) was implemented into the model as the loss and metric while the Adaptive Moment estimation (Adam) optimizer was chosen for the model optimizer to help the model optimize the learning process. A summary of the model is presented in Table 3.

Table 3 Summary of the 2DCNN model

Layer	Output Shape	Number of Parameters
Conv2D	(120, 60,12)	120
MaxPooling2D	(24,12,12)	0
Conv2D	(11,5,6)	654
MaxPooling2D	(2,1,6)	0
Conv2D	(1,1,2)	14
Flatten	2	0
Dense	2	2

3.0 RESULTS AND DISCUSSION

3.1 Comparison Between DL model performance

In this section, all the deep learning models will be compared to each other and they also will be compared with other machine learning models such as Artificial Neural Networks (ANN) and Linear Regression. This comparison helps to determine the most suitable model that is able to perform the gas localization task in the gas map. It will be compared in terms of the accuracy of the model to predict the gas source location and the minimum Euclidian distance error that the model was able to obtain.

The statistical analysis has been conducted for all the generated models by comparing the minimum value, maximum value, mean, standard deviation, and variance for the model Euclidian distance error between the actual gas source and the predicted gas source. Table 4 summarizes the statistical results obtained from each of the models. The lowest average or mean Euclidean distance value was obtained by the 1DCNN model with a value of 0.061 m while the highest mean was obtained by the linear regression model with a value of 0.728 m. This result implies that the deep learning application is better compared to the machine learning algorithm in terms of localizing the gas source. Even though 2DCNN has a larger error compared to other deep learning models the error is still smaller when compared to the machine learning model. The DNN and ANN models also show a promising result when the mean error has a small difference compared to the 1DCNN error and the error is in the acceptable range. The smallest standard deviation also was obtained by 1DCNN which shows less spread of Euclidian distance error compared to 2DCNN and linear regression model.

Table 4 Euclidian distance error statistic comparison between different model

Model Type	Distance range	Euclidian Distance Error (m)			Mean (m)	Standard Deviation (m)	Variance (m^2)
		Average	Min.	Max			
DNN	**(0,0.3]**	0.097	0.000	0.776	0.120	0.109	0.012
	(0.3,0.7]	0.400					
	(0.7,1.0]	0.739					
	>1.0	nan					
1DCNN	**(0,0.3]**	0.062	0.000	0.749	0.061	0.040	0.002
	(0.3,0.7]	0.497					
	(0.7,1.0]	0.731					
	>1.0	nan					
2DCNN	**(0,0.3]**	0.181	0.013	3.572	0.594	0.479	0.230
	(0.3,0.7]	0.482					
	(0.7,1.0]	0.821					
	>1.0	1.500					
ANN	**(0,0.3]**	0.129	0.004	1.561	0.209	0.186	0.283
	(0.3,0.7]	0.395					
	(0.7,1.0]	0.782					
	>1.0	1.561					
Linear Regression	**(0,0.3]**	0.194	0.013	3.382	0.728	0.540	0.291
	(0.3,0.7]	0.480					
	(0.7,1.0]	0.835					
	>1.0	1.492					

To get better visualization for each model error, all the error in Table 4 has been plotted using the box plot as shown in Figure 7.

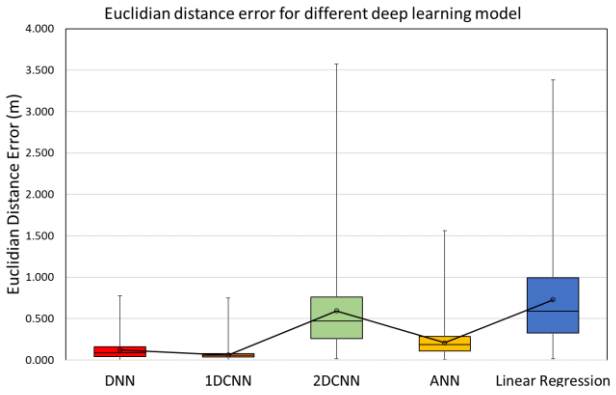


Figure 7 Euclidean distance error whisker plot for a different model

The whisker plot functions to visualize the maximum and minimum error that the models have obtained from the prediction result. The small circle in the box plot indicates the mean error for each model. The box plot shows that the 2DCNN model has the longest whisker for the maximum error and is followed by the linear regression model. However, when looking at the median and mean for the linear regression model it has a larger error compared to the 2DCNN model and other models. The 1DCNN has the smallest data distribution followed by the DNN model and ANN model. This result indicates that all of these models namely DNN, 1DCNN, and ANN have a promising result in predicting the gas source within the gas map. However, 1DCNN shows better accuracy and performance compared to both mentioned models.

Figure 8 shows the performance of each model after being given some random GDM dataset to perform the gas localization prediction. There are four different ranges for the Euclidian distance error between the predicted and actual gas source location. The distance range between 0.0 m to 0.3 m was considered the most accurate, a distance range between 0.3 m to 0.7 m was considered to be moderately accurate, the distance between 0.7 m to 1.0 m was considered to be less accurate, and lastly, distance more than 1.0 m was considered as not accurate. From the figure, it shows that the 1DCNN model gains a higher percentage of occurrence to predict the location of gas sources between the range 0.0 m to 0.3 m with 90.3%. Then it was followed by the DNN model with 89% and the ANN model with 79.3%. This result supports the result from the whisker plot which shows these three models have a high tendency to be considered to predict the location of the gas source. The graph agrees with the previous statistical result. For 2DCNN and linear regression model they have moderate accuracy since most of

the predicted gas location falls between the range of 0.3 m to 0.7 m with 36.5% and 35.9% respectively. However, the linear regression model has the lowest performance compared to 2DCNN and other models because 26.2% of the samples' gas source was predicted to exceed 1 m from the actual location. The graph agrees with the previous statistical result.

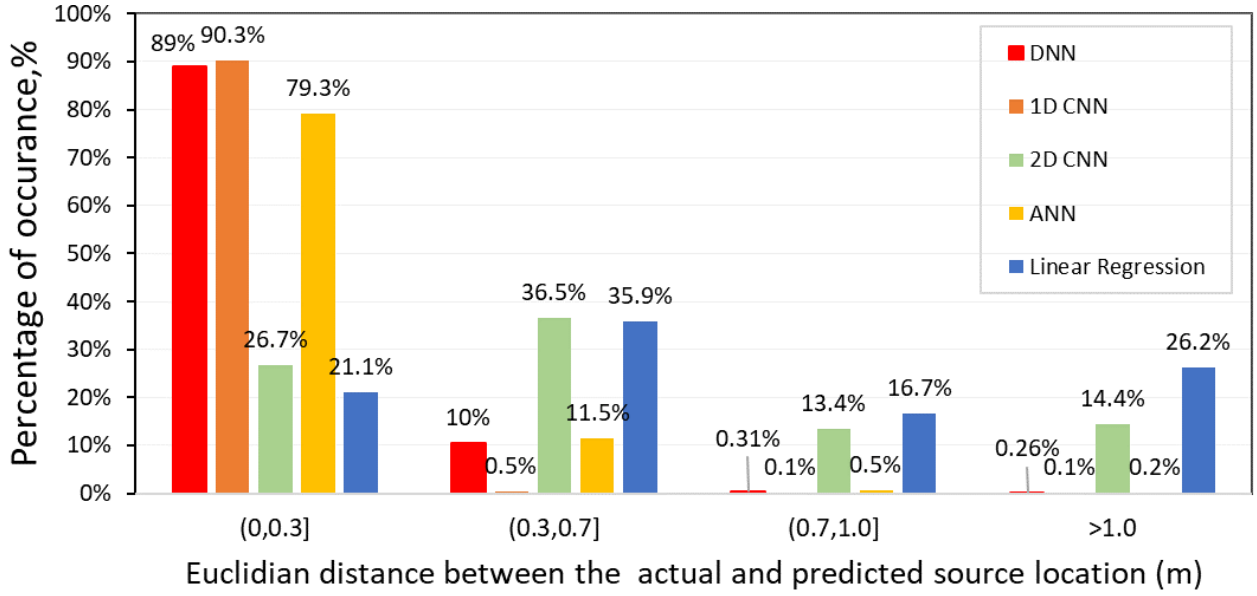
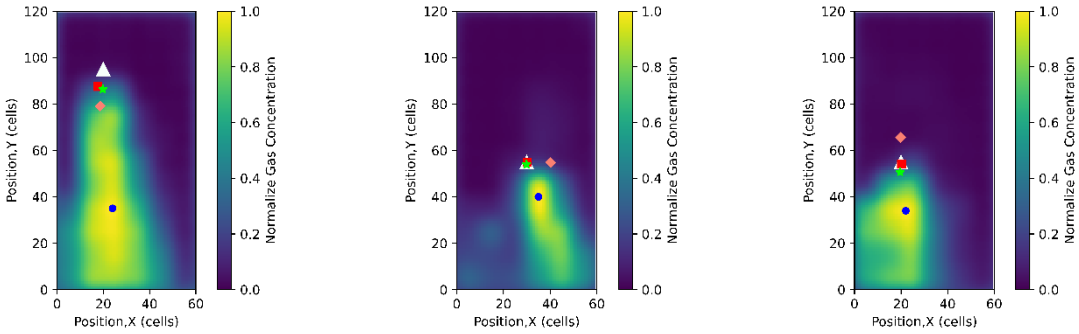


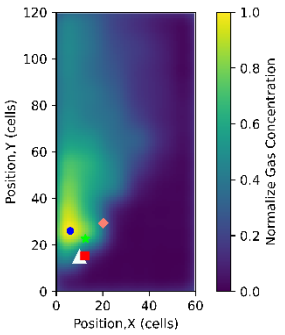
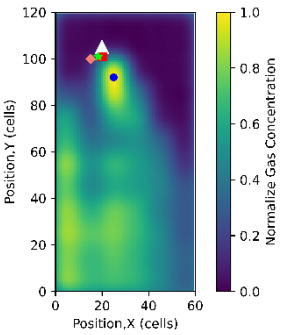
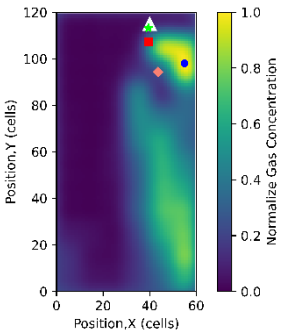
Figure 8 Percentage of occurrence for a different model

Table 5 shows six different random samples of GDM that represent the gas distribution concentration by the yellowish color for clear visualization. The higher the intensity of the yellow color the higher the gas concentration. There is five different shapes and color that represent the actual location of the gas source and the location of gas that is predicted by the different method. The actual gas source is represented by the white triangle while the red square, pink diamond, and green star are represented by gas predicted by the DNN, 1DCNN, and 2DCNN respectively. There is also a blue circle that represents the conventional method namely Global Maximum[20]. This method represents the highest value of concentration in the gas map.

From the table, it can be observed that most of the locations of gas predicted by the Global Maximum method were far away from the actual gas source since the gas plume may move away from the actual gas location. It revealed that the conventional method has a low accuracy since it always assumes the region of high gas concentration has a high tendency to be the gas source. The location of gas predicted by the 1DCNN and DNN was near to the actual gas source and in addition, the locations predicted by both models also were near to each other. However, the location of gas predicted by the 2DCNN model was far from the actual gas source but it is better than the Global Maximum method.

Table 5 Location of predicted gas by different artificial intelligence model

	Actual location, DNN location, 1DCNN, 2DCNN, Global Maxima
GDM	

Actual	(20.0, 95.0)	(30.0, 55.0)	(20.0, 55.0)
DNN	(17.7, 87.5)	(30.1, 55.0)	(20.2, 54.3)
1DCNN	(19.8, 86.3)	(29.9, 53.9)	(19.59, 50.6)
2DCNN	(18.6, 79.1)	(40.2, 54.8)	(19.8, 65.6)
Global. Maximum	(24.0, 35.0)	(35.0, 40.0)	(22.0, 34.0)
GDM			
Actual	(110.0, 15.0)	(120.0, 105.0)	(140.0, 115.0)
DNN	(12.3, 15.3)	(20.1, 100.9)	(39.6, 107.2)
1DCNN	(12.5, 22.7)	(18.3, 100.9)	(39.5, 113.2)
2DCNN	(20.2, 29.3)	(15.1, 99.9)	(43.6, 94.3)
Global Maximum	(6.0, 26.0)	(25.0, 92.0)	(55.0, 98.0)

4.0 CONCLUSION

In conclusion, this research paper introduces a gas source localization system that employs three distinct deep learning techniques, namely DNN, 1DCNN, and 2DCNN. These models are trained using a real-scenario of gas distribution map database to determine the location of gas within the gas map. Notably, the 1DCNN model is observed to outperform the DNN and 2DCNN models with an accuracy rate of 90.3%, compared to 89.0% and 26.7% respectively. It reveals that the 1DCNN effectively captures gas dispersion patterns from array reading and accurately predicts the gas source location compared to the 2DCNN model which is better suited for image data. Furthermore, the findings also suggest that the 1DCNN and DNN models surpass traditional machine learning models which are ANN and linear regression. The accuracy rates of these machine learning models are only 79.3% and 21.1%, respectively. Overall, the 1DCNN model is capable of predicting gas source locations with the highest accuracy, with an Euclidean distance range of 0 m to 0.7 m between the actual and predicted locations.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this paper.

Acknowledgment

The author would like to acknowledge the financial support through the Fundamental Research Grant Scheme (FRGS) under the grant number of FRGS/1/2022/TK08/UNIMAP/02/65 from the Minister of Education Malaysia.

References

[1] C. Bilgera, A. Yamamoto, M. Sawano, H. Matsukura, and H. Ishida. 2018. Application of Convolutional Long Short-term Memory Neural Networks to Signals Collected from a Sensor Network for Autonomous Gas Source Localization in Outdoor Environments. *Sensors (Switzerland)*. 18(12). Doi: 10.3390/s18124484.

[2] F. Rahbar, A. Marjovi, and A. Martinoli. 2019. Design and Performance Evaluation of an Algorithm based on Source Term Estimation for Odor Source Localization. *Sensors (Switzerland)*. 19(3). Doi: 10.3390/s19030656.

[3] C. Liu, Y. Li, L. Fang, and M. Xu. 2019. New Leak-localization Approaches for Gas Pipelines using Acoustic Waves. *Meas. J. Int. Meas. Confed.* 134: 54-65. Doi: 10.1016/j.measurement.2018.10.052.

[4] X. Liu, Z. Peng, X. Liu, and R. Zhou. 2020. Dispersion Characteristics of Hazardous Gas and Exposure Risk Assessment in a Multiroom Building Environment. *Int. J. Environ. Res. Public Health*. 17(1). Doi: 10.3390/ijerph17010199.

[5] N. Evalina and H. A. Azis. 2020. Implementation and Design Gas Leakage Detection System using ATMega8 Microcontroller. *IOP Conf. Ser. Mater. Sci. Eng.* 821(1): 012049. Doi: 10.1088/1757-899X/821/1/012049.

[6] C. Bilgera, A. Yamamoto, M. Sawano, H. Matsukura, and H. Ishida. 2018. Application of Convolutional Long Short-term Memory Neural Networks to Signals Collected from a

Sensor Network for Autonomous Gas Source Localization in Outdoor Environments. *Sensors (Switzerland)*. 18(12). Doi: 10.3390/s18124484.
[7] L. Marques, A. Martins, and A. T. de Almeida. 2005. Environmental Monitoring with Mobile Robots. *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 3624-3629. Doi: 10.1109/IROS.2005.1545133.
[8] X. xing Chen and J. Huang. 2019. Odor Source Localization Algorithms on Mobile Robots: A Review and Future Outlook. *Rob. Auton. Syst.* 112: 123-136. Doi: 10.1016/j.robot.2018.11.014.
[9] B. Bayat, N. Crasta, A. Crespi, A. M. Pascoal, and A. Ijspeert. 2017. Environmental Monitoring using Autonomous Vehicles: A Survey of Recent Searching Techniques. *Curr. Opin. Biotechnol.* 45(645141): 76-84. Doi: 10.1016/j.copbio.2017.01.009.
[10] S. Asadi and A. Lilienthal. 2015. Approaches to Time-Dependent Gas Distribution Modelling. *2015 Eur. Conf. Mob. Robot. ECMR 2015 - Proc.* Doi: 10.1109/ECMR.2015.7324215.
[11] A. Lilienthal and T. Duckett. 2004. Building Gas Concentration Gridmaps with a Mobile Robot. *Rob. Auton. Syst.* 48(1): 3-16. Doi: 10.1016/j.robot.2004.05.002.
[12] A. J. Lilienthal, M. Reggente, M. Trinca, J. L. Blanco, and J. Gonzalez. 2009. A Statistical Approach to Gas Distribution Modelling with Mobile Robots - The Kernel DM+V Algorithm. *2009 IEEE/RSJ Int. Conf. Intell. Robot. Syst. IROS 2009*. 570-576. Doi: 10.1109/IROS.2009.5354304.
[13] M. G. W Khalaf, C. Pace. 2008. Gas Detection via Machine Learning. *Int. J. Comput. Electr. Autom. Control Inf.* 2(1): 61-65.
[14] S. Mahfouz, F. Mourad-chehade, P. Honeine, J. Farah, and H. Snoussi. 2016. Machine Learning in WSNs. 16(14): 5795-5804.
[15] H. Kim, M. Park, C. W. Kim, and D. Shin. 2019. Source Localization for Hazardous Material Release in an Outdoor Chemical Plant via a Combination of LSTM-RNN and CFD Simulation. *Comput. Chem. Eng.* 125: 476-48. Doi: 10.1016/j.compchemeng.2019.03.012.
[16] H. L. Yu, B. H. Chen, K. S. Kim, P. Siwayanan, S. Y. T. Choong, and Z. H. Ban. 2022. Source Localization for Illegal Plastic Burning in Malaysia via CFD-ANN Approach. *Digit. Chem. Eng.* 3(March): 100029. Doi: 10.1016/j.dche.2022.100029.
[17] A. S. A. Yeon, A. Zakaria, S. M. M. S. Zakaria, R. Visvanathan, K. Kamarudin, and L. M. Kamarudin. 2022. Gas Source Localization via Mobile Robot with Gas Distribution Mapping and Deep Neural Network. November: 120-124. Doi: 10.1109/ice3is56585.2022.10010251.
[18] A. Jain *et al.* 2020. Overview and Importance of Data Quality for Machine Learning Tasks. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 3561-3562. Doi: 10.1145/3394486.3406477.
[19] S. Syed Zakaria *et al.* 2015. Development of a Scalable Testbed for Mobile Olfaction Verification. *Sensors*. 15(12): 30894-30912. Doi: 10.3390/s151229834.
[20] K. Kamarudin *et al.* 2018. Integrating SLAM and Gas Distribution Mapping (SLAM-GDM) for Real-time Gas Source Localization. *Adv. Robot.* 32(17): 903-917. Doi: 10.1080/01691864.2018.1516568.
[21] J. G. Monroy, J. L. Blanco, and J. Gonzalez-Jimenez. 2016. Time-variant Gas Distribution Mapping with Obstacle Information. *Auton. Robots*. 40(1): 1-16. Doi: 10.1007/s10514-015-9437-0.
[22] Y. LeCun, Y. Bengio, and G. Hinton. 2015. Deep Learning. *Nature*. 521(7553): 436-444. Doi: 10.1038/nature14539.