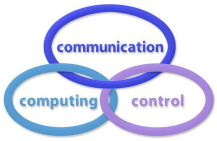
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 18, Issue: 3, Month: June, Year: 2023
Article Number: 5084, https://doi.org/10.15837/ijccc.2023.3.5084



CCC Publications

Application of Deep Neural Network for Gas Source Localization in an Indoor Environment

Z.H.M. Juffry, K. Kamarudin, A.H. Adom, M.F. Miskon, L.M. Kamarudin
A. Zakaria, S.M. Mamduh, A.N. Abdullah

Zaffry Hadi Mohd Juffry, Kamarulzaman Kamarudin*, Abdul Hamid Adom, Abdulnasser Nabil Abdullah
Faculty of Electrical Engineering and Technology
Universiti Malaysia Perlis, Malaysia,
02600 Arau Perlis, Malaysia
*Corresponding author: kamarulzaman@unimap.edu.my
zaffryhadi@studentmail.unimap.edu.my, abdhamid@unimap.edu.my, abdulnasser@unimap.edu.my

Muhammad Fahmi Miskon
Faculty of Electrical Engineering
Universiti Teknikal Malaysia Melaka,
76100 Durian Tunggal, Melaka, Malaysia
fahmimiskon@utem.edu.my

Latifah Munirah Kamarudin
Faculty of Electronic Engineering and Technology
Universiti Malaysia Perlis, Malaysia
02600 Arau Perlis, Malaysia
latifahmunirah@unimap.edu.my

Ammar Zakaria, Syed Muhammad Mamduh Syed Zakaria
Center of Excellence for Advance Sensor Technology (CEASTech)
Universiti Malaysia Perlis, Malaysia,
02600 Arau Perlis, Malaysia
ammarzakaria@unimap.edu.my, smmamduh@unimap.edu.my

Abstract

Nowadays, the quality of air in the environment has been impacted by the industry. It is important to make sure our ambient air especially in an indoor environment is clean from contaminating particles or harmful gases. Therefore, the air quality inside the indoor environment should be monitored regularly. One of the major problems, when a particular environment has been contaminated by harmful gases, is finding the source of the emission. If the indoor environment has been contaminated by a harmful source it should be instantly localized and eliminated to prevent severe casualties. In this paper, we propose the utilization of synthetic data generated by the Computational Fluid Dynamic (CFD) approach to train the Deep Neural Network (DNN) model called CFD-DNN to perform gas source localization in an indoor environment. The model is

capable to locate the contaminated source within a small area of an indoor environment. A total of 361 datasets with different locations of contaminated source release have been obtained using the CFD approach. The obtained dataset was divided into training and testing datasets. The training dataset was used for the model training process while the testing dataset is fed into the model to test model reliability to predict the gas source location. The Euclidian distance equation was used to measure the distance error between the actual and predicted location of the source. The result shows that the model is capable to locate the gas source within a minimum and maximum error of 0.03m to 0.46m respectively.

Keywords: gas source localization, deep neural network, computational fluid dynamic, harmful gases, deep learning.

1 Introduction

Gas source localization (GSL) commonly is known to be a task to locate a harmful gas location in a particular contaminated area [1]. Researchers in the harmful gas field have developed various algorithms to find the gas source location. The motivation is due to there being many cases related to the harmful gas accidents reported lately. For example, in 2016, an incident of harmful gas leakage occurred at a fertilizer company in the port city of Chittagong, Bangladesh. No deaths were reported but 250 people had fallen ill due to toxic ammonia inhalation [2]. There is much more gas leakage incident reported by the media lately. In Bangladesh, from the report data, the accident of gas leakage jumped from 3,447 in 2016 to 4,428 in 2019 [3]. It is important to find the gas source leakage as soon as possible to prevent more severe damage. The rescuers and the firefighters need to enter the contaminated area to find the source of harmful gas release by using complete personal protective equipment (PPE). This conventional method is too dangerous and risks their lives if there is an explosion of gas, inhalation, or exposure to a high level of harmful gasses. Some of the areas are likely to release harmful gasses such as landfill sites [4], mines [5], and factories [6]. Normally, stationary gas sensors are used inside the factory to monitor the leakage of the gas but they are not for gas localization purposes.

There are two methods to perform the GSL which are by utilizing the mobile robot technology that is equipped with a gas sensor [7] or the conventional method that uses the stationary gas sensor [8]. Both of these methods have their pros and cons. The advantage of using the mobile robot to perform the GSL is the robot is easier to deploy and being able to enter the contaminated area that cannot be reached by human beings. It is also able to do repetitive tasks without getting exhausted. However, by considering the condition of the real environment the gas concentration distribution tends to be patchy, intermittent, and time-variant [9]. This led to the misinterpretation of data collected by the robot to locate the gas source. Otherwise, the stationary gas sensor is able to collect particular gas concentration information in a particular area. The advantage is that it is able to provide richer information about the harmful gasses in a particular environment and is able to collect important data regularly. However, the disadvantage of a stationary sensor is it will consume a lot of time to be installed and is not flexible, it is only able to monitor a certain "target" area.

The utilization of robots to find the gas source is known to be a mobile robot olfaction (MRO) [10]. Researchers in the MRO field have tried many approaches to make a robot able to find the gas source location. This includes mimicking the animals' behavior to find their mates and food. This is known to be a bio-inspired approach since it is inspired by the behavior of living organisms such as silkworms, lobsters, and beetle. The study of this approach was started by [11]. This approach is divided into two categories which are anemotaxis and chemotaxis. The chemotaxis strategy is based on the detection of concentration differences between two different positions or two chemical sensors. In simple words, the concentration of the gas will drive the robot mechanism toward the gas sources [12]. Anemotaxis instead refers to a mechanism in which the movement of an organism (or mobile robot) is determined by the perceived airflow [13]. Follow-up wind direction can be a theoretical approach that brings the mobile robot toward the gas sources which are motivated by the moth's plume tracking capability. Both of these approach categories are able to work fine in the indoor environment that has a consistent gas distribution instead of the outdoor environment since the strong movement of wind causes the gas concentration to fluctuate and is not consistent. Because of this problem the researcher tried to

explore another approach that treats the gas concentration as the probabilistic way to find the gas source location.

The probabilistic approach in MRO was proposed by [14]. They proposed a method called the infotaxis approach which is also known to be a Bayes inference method. This method treats the gas measurement as a probabilistic distribution. The robot will maneuver toward the gas area based on the minimum entropy criterion such that it either stays at the same location or moves to another neighborhood. It has a similarity with the nature of stationary sensors which try to predict the gas distribution in the environment but the difference is the gas sensor is dynamic and is carried by the mobile robot. Then, in the probabilistic approach, there is another method called the gas distribution map (GDM). It is a process of creating a representation of how gas disperses in the environment from a set of temporally and spatially distributed measurements of relevant gas data [15], [16]. The advantage of using GDM is that the robot is able to achieve measurement not only in a particular area but also in other areas and still be able to provide the source of the gas. Usually, the location of the single or multiple gas sources in the GDM was predicted by global minimum or local minima respectively.

In order to accurately represent air transport and dispersion in extremely complex systems, Computational Fluid Dynamic (CFD) methodologies are now used. The utilization of CFD to solve the governing set of equations numerically to perform the dispersion of harmful gas simulation has become more and more popular in the MRO field. The reason is that in many realistic conditions gas is dispersed by turbulent advection. The turbulent flow causes the gas plume to follow chaotic trajectories. It will result in fluctuating, intermittent patches of the gas concentration.[17] has used the CFD method to predict the indoor contaminant plume for testing three-dimensional (3D) and two-dimensional (2D) plume search algorithms. The study reported that the 3D search algorithm obtains better results compared to the 2D search algorithm. There is also another researcher who utilizes the CFD to generate airflow and gas concentration and then imports it to other software such as GADEN to simulate the olfaction mobile robot algorithm [18]. In this work, the gas concentration synthetic data will be generated by the CFD simulation to provide the data for the model's training datasets. It will support the gas localization assessment without the use of MRO. Instead, we used the Deep Neural Network (DNN) algorithm to model and predict the location of gas source release inside a small empty room. There are several studies that have focused on the utilization of artificial intelligence for gas localization assessment which are using machine learning such as Support vector machines (SVMs) [19] and kernel ridge regression [20]. Then, the recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM-RNN) and Feedforward Neural Network (FNN) model used in training for the identification of the leak source location has been studied by [21]. 460 scenarios have been generated by using CFD simulation and fed into two neural network models. They find that, after comparing the accuracy of both models for the location assessment of leak spots, LSTM-RNN accuracy is around 20% higher compared to the FNN model. Besides, the combination of Convolutional Neural Network (CNN) and Long-short term memory (LSTM) also has been performed to estimate the location of gas in an outdoor environment [22]. The study used an array of the gas sensor in the real environment to generate the sequential dataset to train the model and they found that artificial neural network (ANN) is a promising prospect for the GSL tasks. In [23] they tried to predict the plastic burning location by using an ANN model that includes 16 inputs,4 hidden, and 12 output neurons. They generated the data such as burning location, wind speeds, and wind direction using the CFD simulation and then trained the ANN model. They claim that the trained model is able to achieve up to 85.71% validity with an average error of 3.86%. There is a study that focuses on the utilization of DNN for localizing gas sources based on the Gas Distribution Map (GDM)[24]. They collected the dataset through four different GDM samples and performed the augmentation process to increase the number of datasets from the original GDM samples. Then, the dataset was fed into the DNN model which has four hidden layers. The study found that the DNN model was able to predict the gas source with a maximum distance error of 1.15m. However, the study does not focus to optimize the DNN model hyperparameter to get the best model. This paper will focus on the synthetic dataset of gas dispersion using the CFD method without being required to perform the data augmentation process to increase the number of datasets. Then the dataset fed into the DNN model and the configuration

of different model hyperparameters will be tested to find the best model performance.

The remaining sections of this paper are arranged as follows: Section 2 introduces the method for the development of the synthetic datasets using CFD in order to perform the model training process as well as the approach to develop the DNN algorithm model to predict the gas source location. Section 3 will cover the discussion about the result of the training process as well as the performance of the generated DNN model. Then, plans for future research and a summary of the analysis are included in the final section.

2 Methodology

This section will explain the experimental setup to perform the harmful gas emission in an indoor environment using the simulation method. The setting of the model boundary condition, the location of the gas source, size of the indoor environment will be explained in this section. This section also will illustrate the method of the DNN model development to train and test the synthetic dataset from the simulation.

2.1 Simulation model for data collection

The study is performed by using the computational geometry modelled based on the small size of an office room. The ANSYS/Workbench 2020 R2 was used to perform the ANSYS/Fluid Flow (CFX) analysis to simulate the harmful gas dispersion inside the room. ANSYS/CFX is known to be a high-performance computational fluid dynamic (CFD) software recognized for its robustness, high computation accuracy, and speed for gas analysis. In CFD, the basic equations that govern the conservation of mass, momentum, and energy balance are solved numerically for a given initial, flow domain, and boundary conditions. The flow of a Newtonian fluid is governed by the continuity and the momentum conservation equations for total energy flow, described by equation (1) and equation (2) respectively[25];

Continuity:

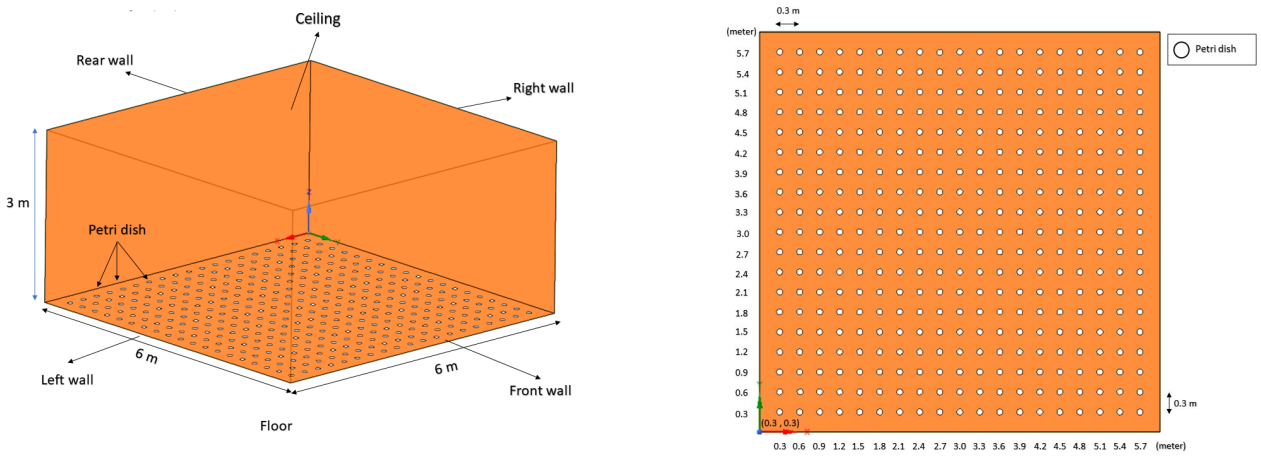
$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u_i)}{\partial x_i} = 0 \tag{1}$$

Momentum balance:

$$\frac{\partial \rho u_i}{\partial t} + \frac{\partial (\rho u_i u_j)}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j}\left[\mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) + \lambda \frac{\partial u_k}{\partial x_k}\right] + \rho g_i \tag{2}$$

Here, ρ is the density, u_i is the velocity component in the ith direction, p is the pressure, μ is the viscosity of the medium, λ is the second coefficient viscosity (the term involving this variable is assumed to be negligible) and g_i is the component of gravitational acceleration in the ith direction. For equations details see the ANSYS CFX-Solver Theory Guide [26].

ANSYS/SpaceClaim is a three-dimensional(3D) modelling software used to generate the model geometry. The dimension of the simulated office room is 6m x 6m x 3m (length x width x height) and does not contain any ventilation system to prevent the gas plume from being affected by the strong wind movement. The room consists of four walls, a ceiling, and a flat surface floor. There are no windows, doors, or obstacles such as the office chair and table inside the room to create a simple environment and to eliminate the obstacle effect on the gas dispersion. The ethanol vapor was released into the domain through a simulated Petri dish with a size of 5 cm in diameter. The gas was released at 361 different locations inside the room. Thus, a total of 361 sets of CFD simulation runs were conducted in this work. All the Petri dishes arrange into a 19 x 19 matrix and have their own coordinates, namely on the x-axis and y-axis. The distance between each of the Petri dishes is 0.3m and they are placed on the ground level to cover the office floor area. (Figure 1) shows the office geometry model with the location of each of the gas source's release points from the top view.



(a) (b)

Figure 1: (a) Isometric view of model geometry (b) Petri dish position for gas inlet

2.2 Mesh generation

A tetrahedron mesh type with a size of 0.2m was generated for the domain model using meshing software which is ANSYS/ICEM. The total number of mesh elements and mesh nodes generated after the meshing process is 116,058 and 21,319 respectively. (Figure 2) shows the generated mesh for the simulated geometry room.

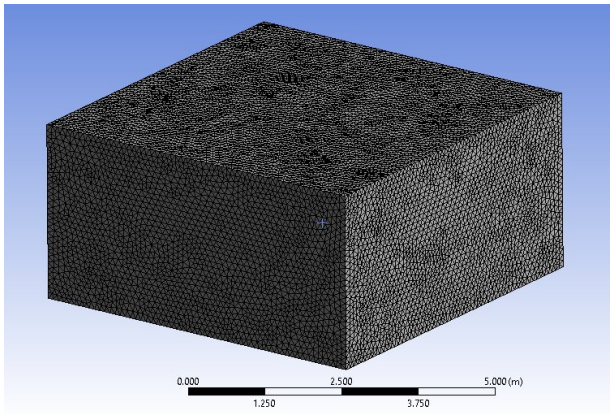


Figure 2: Model geometry mesh generation with a tetrahedron type

2.3 Boundary setting

Since the wind velocities in the present study are fairly low and no heat transfer is considered, the walls, ceiling, and floor will be set as the no special, wall-induced effect boundary condition and the heat transfer between them is neglected. Then, each of the simulated Petri dishes is set to be inlet boundary condition. The ethanol vapor was released at a rate of 0.01kg/s from the inlet.

The simulations were performed on Intel i5 core computer generation. The flow type is total temperature and steady state. The ambient pressure and temperature are set up constant to be 1 atm and 25°C respectively. For the global initialization, the turbulent viscosity μ_t was calculated by the k-epsilon (k-ϵ) turbulence closure model defined as follows:

$$\mu_t = C_\mu \frac{k^2}{\epsilon} \tag{3}$$

$$\frac{\partial}{\partial t}(\rho k) + \frac{\partial}{\partial x_j}[u_j(\rho k)] = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_k}\right)\frac{\partial k}{\partial x_j}\right] + G_k - \rho\epsilon \tag{4}$$

$$\frac{\partial}{\partial t}(\rho\epsilon) + \frac{\partial}{\partial x_j}[u_j(\rho\epsilon)] = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_t}\right)\frac{\partial t}{\partial x_j}\right] + c_{1\epsilon}\frac{\epsilon}{k}G_k - c_{2\epsilon}\rho\frac{\epsilon^2}{k} \tag{5}$$

In equation (3) – (5), the empirical constants are standard for the (k-ϵ) turbulence model which have the following defaults values: $C_\mu = 0.09$, $c_{1\epsilon} = 1.44$, $c_{2\epsilon} = 1.92$, $\sigma_t = 1.3$, $\sigma_k = 1.0$ [27]. These values are determined from experimental data and are used to model the effects of turbulence on fluid flow. The k-epsilon (k-ϵ) turbulence model is widely utilized in computational fluid dynamics (CFD) to replicate mean flow features in turbulent flow circumstances. It was chosen because of its relative simplicity and it was used by the previous author [14]. The turbulence kinetic energy is denoted as k, which is the measurement of velocity fluctuations variance and has units of m^2/s^2. The turbulence eddy dissipation is denoted as epsilon, ϵ which is the rate at which the velocity fluctuations dissipate and has units of m^2/s^3. Typically, the computational time for each simulation of gas release took about 45 minutes. The total time needed to perform for all locations of gas takes about 1 month.

2.4 Data extraction

After the simulation was done, the result was analysed using ANSYS/CFD-POST application. A horizontal plane is located at the z-axis of the room model with a height of 0.05m from the ground to acquire the ethanol mass fraction contour plot. (Figure 3) below shows several examples of the simulation result for ethanol vapor release at different Petri dish locations (0.3m, 5.7m), (1.2m, 1.5m), (2.7m, 0.3m) and (4.8m, 0.9m) under similar surrounding conditions. Through the contour plot view of the ethanol vapor mass fraction, it can be observed clearly the spread area of the ethanol vapor inside the office room. However, it was found that they have different shapes of dispersion. This is because some of the release points are near to the room's wall causing the ethanol vapor to accumulate on the edge of the wall. The concentration of the ethanol vapor is highest at the release point and decreases as they get far from the release point. Therefore, it can conclude that the CFD simulation generated is valid and agreed well with the theoretical model of gas dispersion. All the data of the ethanol mass fraction on the plane is collected and exported to Excel files. There are 19788 total numbers of nodes on the plane. Each of the nodes contains an ethanol mass fraction reading. The total Excel data generated is 361 files by referring to 361 different locations of the gas.

2.5 Deep Neural Network (DNN)

In this work, a deep neural network (DNN); more specifically, a multilayer perceptron (MLP) will be used to predict the location of a gas source in an office room according to the collected dataset generated by CFD result. DNN contains several numbers of neurons organized into layers, connected to each other by synaptic weights. It also contains more than one layer of hidden layer units between its input and output. This characteristic allows DNN to solve complex problems. In this study, the DNN will be treated as supervised learning, and the location of the gas source as a linear regression in the x-axis and y-axis direction. By referring to (Figure 4) below there are two phases for the DNN to predict the gas location. The first step is to split the dataset into two groups which are the training and also the testing dataset. The DNN model needs to be trained using the training dataset and after it gets the solid model then it will be used to predict the location of gas accordingly to what it learned from the testing dataset.

2.6 Model generation

The entire process to generate a complete DNN model is implemented by using the TensorFlow package and Python programming code on the Google Collaboratory integrated development environment (IDE). The DNN model is composed of three main layers: the input layer, the hidden layer, and also output layer. The input layer consists in a number of 19787 nodes which refers to the number of nodes on the z-plane in the CFD simulation. Initially, the hidden layer consists of 2 layers and each layer has a number of 300 nodes respectively. The Rectified Linear Unit (ReLU) activation function is

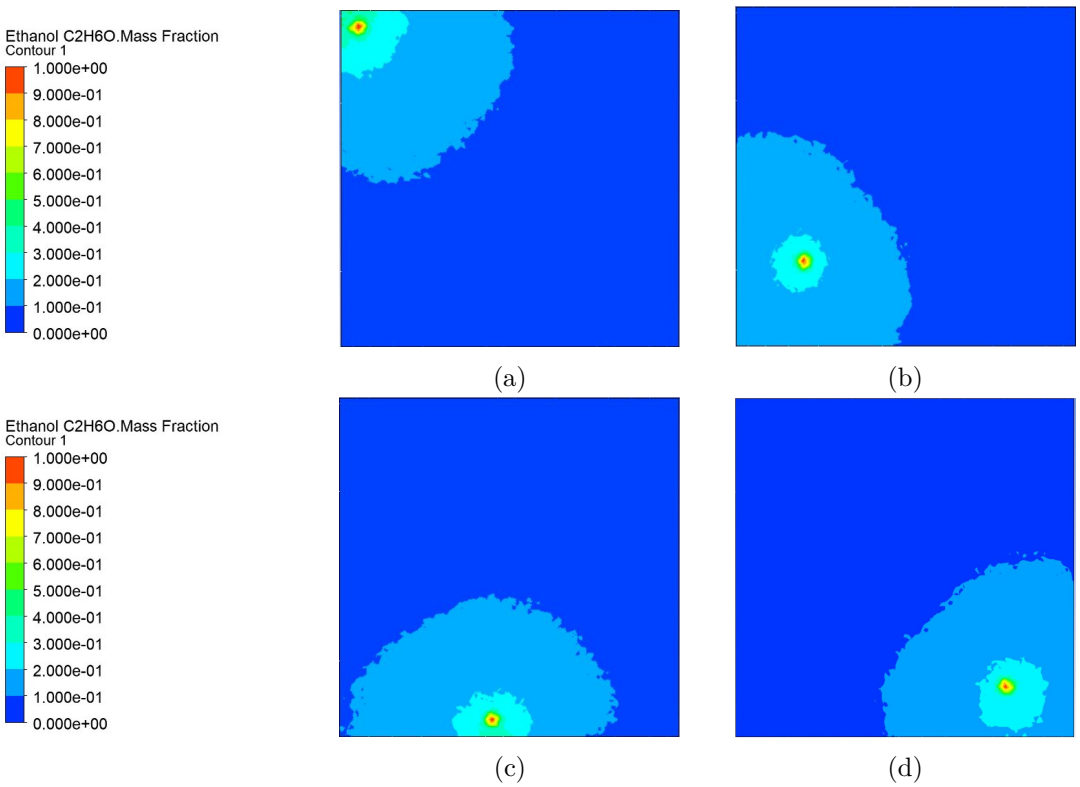


Figure 3: Contour view of ethanol mass fraction concentration (a) location of release point at (0.3m, 5.7m) (b) location of release point at (1.2m, 1.5m) (c) location of release point at (2.7m, 0.3m) (d) location of release point at (4.8m, 0.9m)

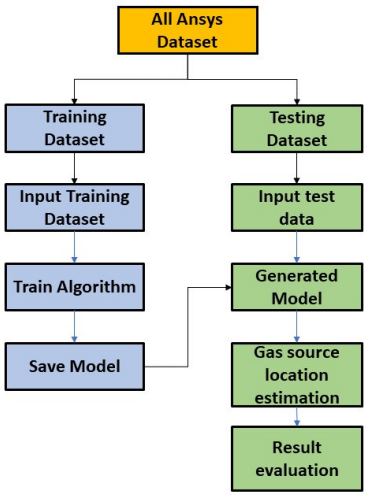


Figure 4: Flowchart of the proposed method

used for the hidden layer since the dataset contains only positive values. The output layer only consists of 2 nodes which refer to the X-coordinate and Y-coordinate for the gas location. The activation function used for the output layer is a linear activation function that treats the output as a regression. The entire DNN model architecture is shown in (Figure 5).

2.7 Model training

Each of the datasets is stacked into an array followed by its two labels which are in term of X-coordinate and Y-coordinate labels. The labels represent the location of simulated Petri dishes that

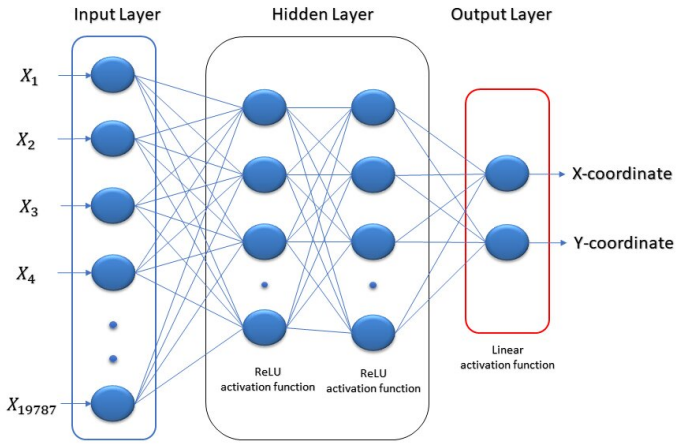


Figure 5: Deep Neural Network (DNN) model architecture

release the ethanol vapor. Then, the array that contains all the training dataset with their labels were randomly fed into the DNN model for the learning process. By entering the input layer with the reading of ethanol mass fraction from the z-plane, it will be performing the multiplying between the first layer node, x_i and synaptic weights, ω_i connected to the first hidden layer to get the result for the first hidden layer node. 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:

$$u = \sum_{i-n}^{n} x_i \omega_i \tag{6}$$

where x_i is

$$x_i = X_n \tag{7}$$

where X_n is the node reading on the z-plane that contained the ethanol mass fraction values. Then,u value is inserted into the linear activation function which is Rectified Linear Unit (ReLU) to define all negative values of u to be 0 or otherwise to be a positive value of u itself.

$$f(u) = \begin{cases} u & \text{for} \quad u \geq 0 \\ 0 & \text{for} \quad u \leq 0 \end{cases} \tag{8}$$

where 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 characteristic of the data remain without disappearing to the output layer. The same process goes onto the second layer of the hidden layer. The weight and bias are updated based on the loss rate function which is Mean Squared Error (MSE) between the actual and predicted gas location until an optimal value is achieved. The MSE equations are as below:

$$MSE_x = \frac{1}{n} \sum_{i=n}^{n} (x - \hat{x})^2 \tag{9}$$

$$MSE_y = \frac{1}{n} \sum_{i=n}^{n} (y - \hat{y})^2 \tag{10}$$

where, n is the number of target data while x and y are the actual locations of the gas source coordinate, $\hat{x}$ and $\hat{y}$ is the predictive location of the gas source. For the purpose to optimize the

learning process, Adaptive Learning Rate (ADAM) was used in this study because it is effective for dealing with nonlinear problems, including outliers while the epoch for maximum learning was set to 300.

2.8 Model evaluation

To evaluate the model performance, the dataset has been divided into two categories which are training and testing set through random selection. From the total dataset, 77% numbers of the data are used for training while 33% numbers of the remaining data are used for testing the model. Then, 30% of the testing data is randomly selected for model validation. This is to avoid over-fitting a particular dataset.

3 Result and Discussion

This section will discuss the result from the generated DNN model in terms of its learning process, the accuracy of the model, and the reliability of the model to predict the gas source location. The distance error between the actual gas source location and the predicted gas source location is calculated using the Euclidean distance formula:

$$d = \sqrt{(x - \hat{x})^2 + (y - \hat{y})^2} \tag{11}$$

where d is the distance, x and y are the locations of the actual gas source. $\hat{x}$ and $\hat{y}$ are the location of the predicted gas source.

3.1 Variable number of hidden layer nodes

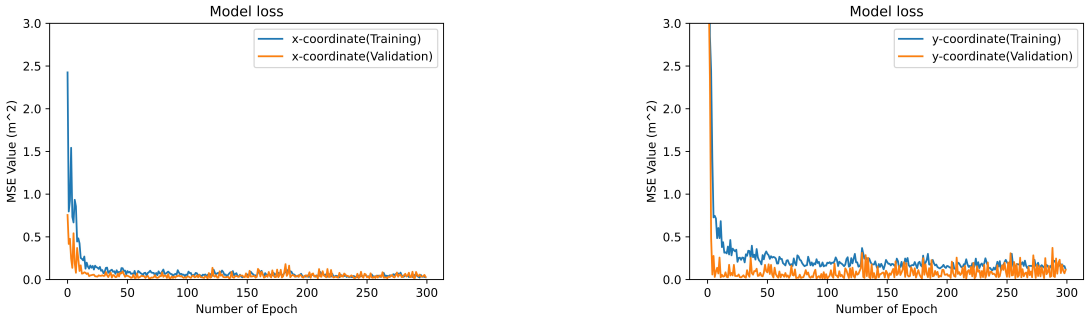
This study also performed some experiments to see the improvement of the model's accuracy to predict gas source location by changing the hyperparameter of the model which is the number of nodes in the hidden layer. There are five different configurations as shown in (Table 1) below. All the configurations consisting of only 2 hidden layers and the number of nodes were chosen heuristically. The best configuration that achieved the best performance has been chosen.

Table 1: Different configuration of the model's hidden layer hyperparameter

Configuration	Number of Nodes	
	1st Hidden Layer	2nd Hidden Layer
1	300	300
2	500	300
3	500	500
4	500	1000
5	1000	1000

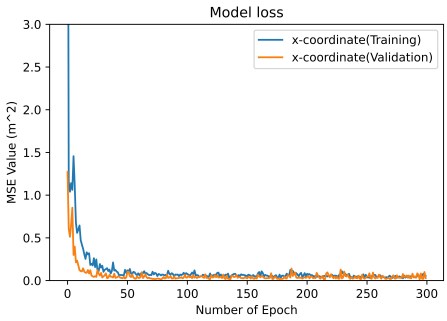
(Figure 6) below shows all the loss graphs for the training and validation of the different models with a different number of nodes in the hidden layer to illustrate the learning performance of the proposed model with a different configuration. The number of model training epochs that refer to the number of training cycles through the dataset is shown on the x-axis while the y-axis shows the MSE loss value of the model for the X-coordinate and Y-coordinate output. Both the training and validating loss graphs for X-coordinate and Y-coordinate output curves follow the downward trend as the number of epochs increases. As the number of dataset epochs increases the mean square error value decreases. This is a good sign that indicates the model was able to learn from the CFD dataset and was able to converge to a certain level of accuracy.

(Table 2) shows the evaluation result of each different model configuration. From the displayed table, the highest model loss error gain after the learning process was from configuration number 5

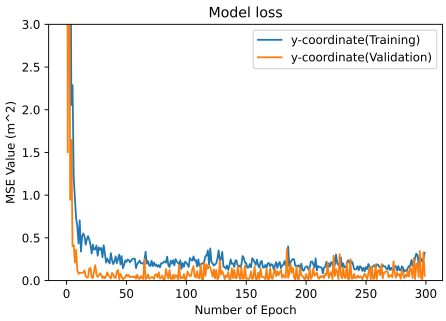


(a)

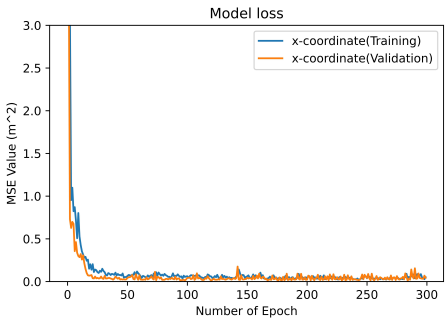
(b)



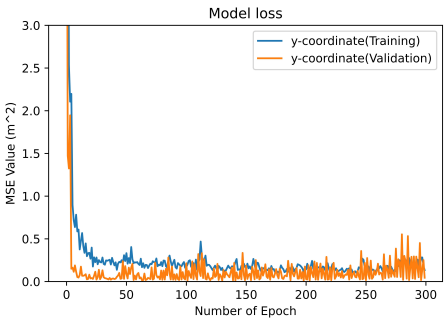
(c)



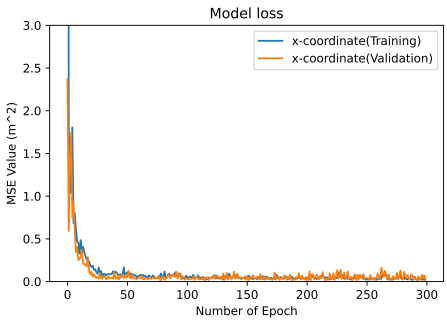
(d)



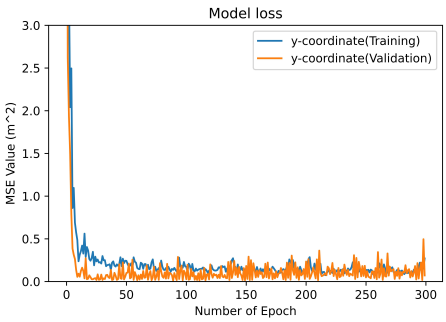
(e)



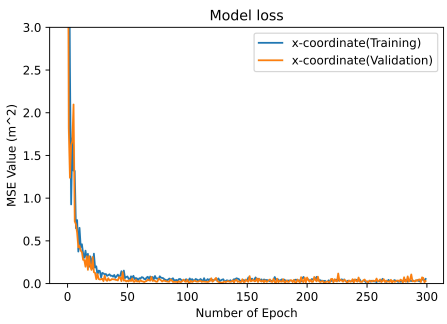
(f)



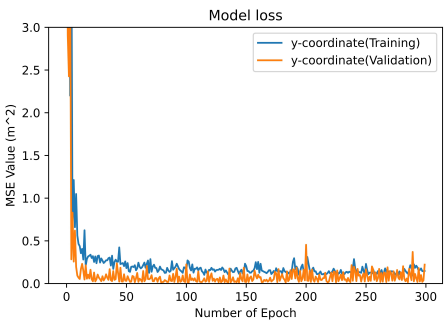
(g)



(h)



(i)



(j)

Figure 6: Training model loss graph (a-b) for configuration 1; (c-d) for configuration 2 ; (e-f) for configuration 3; (g-h) for configuration 4; (i-j) for configuration 5

while the lowest total model error was gained by configuration 3. From this result, configuration 3 will be considered to be the selected model since it has the lowest model loss error. However, relying only on the model evaluation result without considering each model performance to predict the gas source is insufficient. Therefore, each model's performance has been tested by looking at the percentage of occurrence for the model to predict the location of the gas source.

Table 2: Model loss evaluation result after the training process

Configuration	Number of Nodes		Model Loss (m)	X-coordinate loss (m)	Y-coordinate loss (m)
	1st Hidden Layer	2nd Hidden Layer			
1	300	300	0.377	0.036	0.342
2	500	300	0.221	0.033	0.189
3	500	500	0.214	0.041	0.173
4	500	1000	0.279	0.029	0.250
5	1000	1000	0.470	0.033	0.437

(Figure 7) shows the result of all model performances to predict the location of the gas source in terms of the percentage of occurrence of the model to perform the gas source prediction. It represents the percentage of occurrence in 278 data for the training dataset and 83 data for the testing dataset. The model was fed with the data from the training and testing dataset to see whether the model is able to accurately predict the location of the gas source. The Euclidean distance between the gas source location predicted by the DNN model and the actual gas source location was used to evaluate whether the location of the gas source is estimated correctly or not. From the histogram chart, the range of 0.0m to 0.3m was considered as the most accurate distance of the predicted gas location with the true gas location. The distance range between 0.3m to 0.7 was considered as the medium accuracy while above 0.7m was considered as low accuracy for the model.

(Figure (7b) shows that the DNN model with configuration 2 has the highest percentage of occurrence to predict the location of the gas source from the training dataset in the range of 0 m to 0.3m with 73.7% while 24.2% fall within the range between 0.3 to 0.7 and the remaining was fallen between the range of 0.7m and 1.0m which consider to be acceptable to identify the location of the gas source. Note that the number of hidden layers of configuration 2 for the first and second hidden layers was 500 and 300 nodes respectively.

The lowest performance of the model is shown in which is for Configuration 5 which only obtained 17.2% of prediction of the output which falls in the range of 0.0m to 0.3m. Most of the prediction of the output falls between the range of 0.3m to 0.7m with 76.8% of the percentage of occurrence while the remaining data falls in between a range of 0.7m to 1.0m. Configuration 5 has the highest number of nodes at the hidden layers which are 1000 nodes for both hidden layers. This evidence shows that a higher number of nodes in the hidden layer does not provide higher learning and prediction performance for the model. The higher number of nodes in the hidden layer will lead to the increase in the computation time for the model to learn a particular dataset and increase the tendency to be an overfitting model.

(Figure 8) shows the box plot for model prediction error for different model configurations in terms of Euclidean distance between the predicted gas location and true location. The "x" marker shows the Euclidean distance mean error for the output data. The box plot also shows the maximum and minimum error obtained by each of the models and the maximum distance error was obtained by Configuration 5 which is 0.9m shown in (Figure 8e). The majority of the result shows that the mean distance error predicted from the training dataset were much smaller compared to the mean distance predicted using the testing dataset.

For easier to make the comparison between the model to obtain the best model to predict the location of gas in an indoor environment (Figure 9) and (Figure 10) shows the performance of each

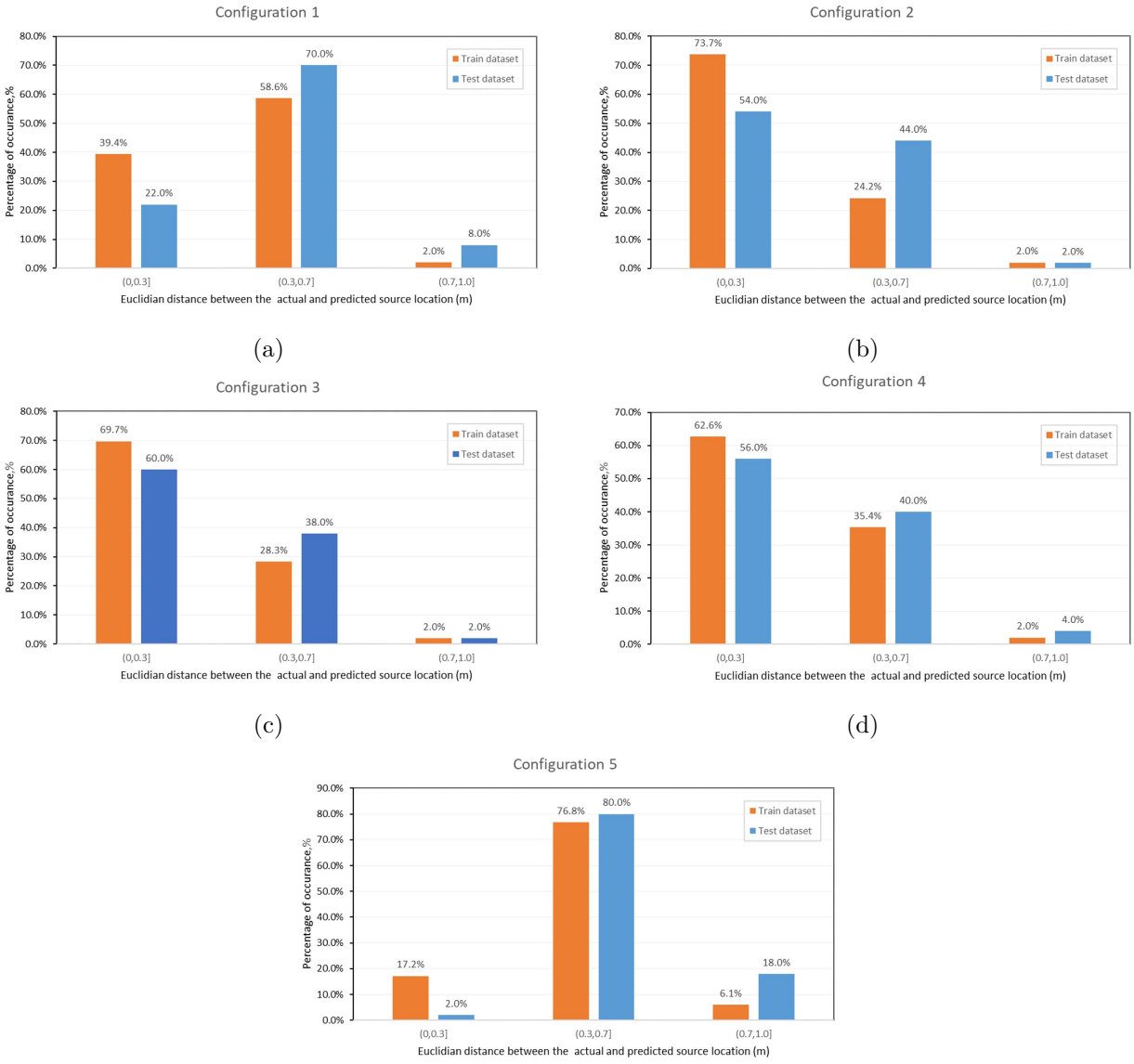


Figure 7: Performance of each DNN model to predict the location of gas source (a) Configuration 1 (b) Configuration 2; (c) Configuration 3; (d) Configuration 4; (e) Configuration 5

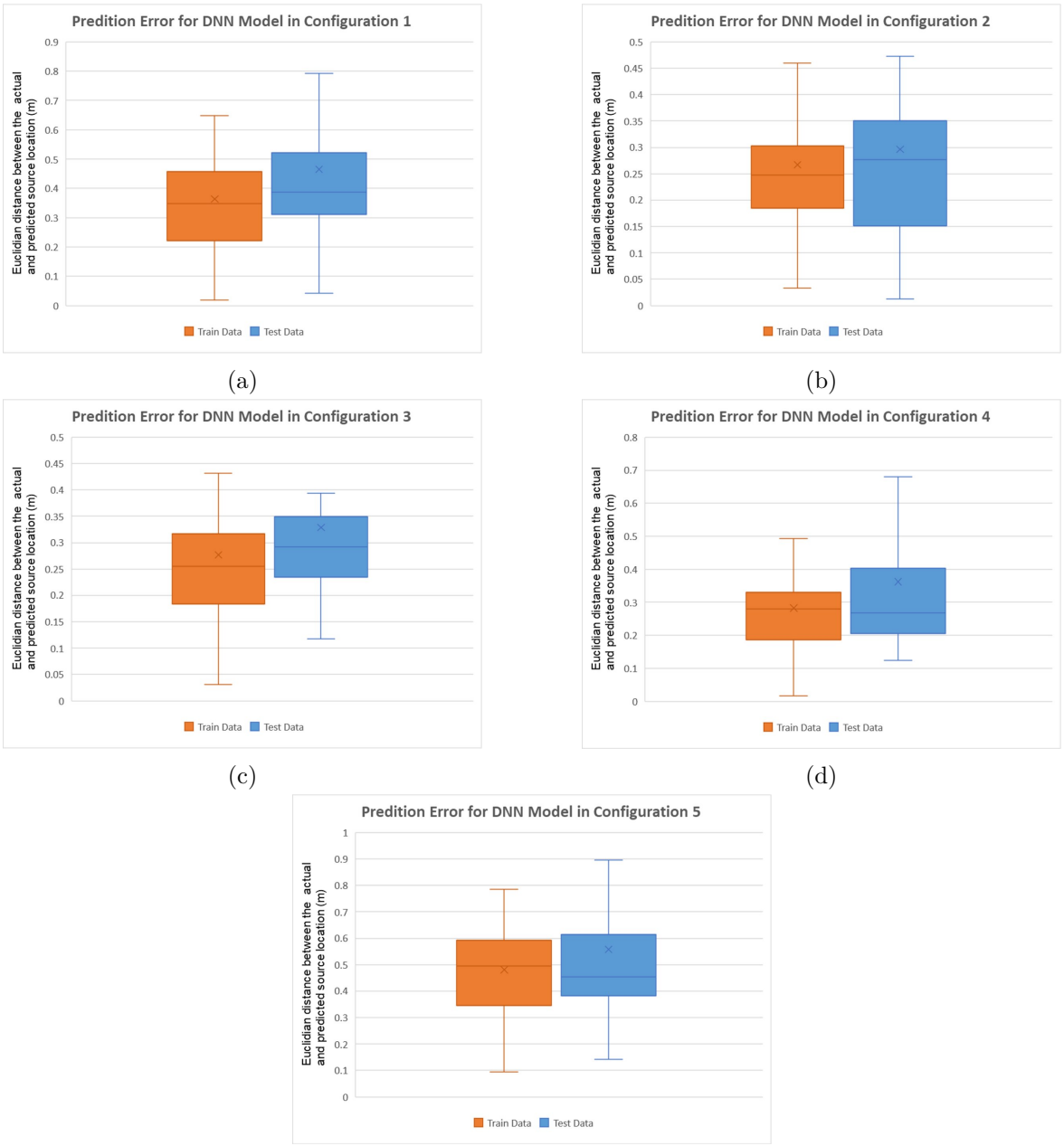


Figure 8: Euclidean distance error for each model (a) Configuration 1 (b) Configuration 2 (c) Configuration 3; (d) Configuration 4; (e) Configuration 5

model in terms of percentage of occurrence and prediction error side by side. From (Figure 9) it is very obvious that the DNN model with configuration 2 has the highest performance to predict the gas source compared to another model configuration. Configuration 2 performance was 56.5% higher compared to the model performance of configuration 5. (Figure 10) also agrees with the result in (Figure 9) which shows that configuration 2 has the lowest Euclidian distance mean error compared to another model configuration. Therefore, from the obtained results, the DNN model with configuration 2 was chosen as the best model to predict the gas source in this study.

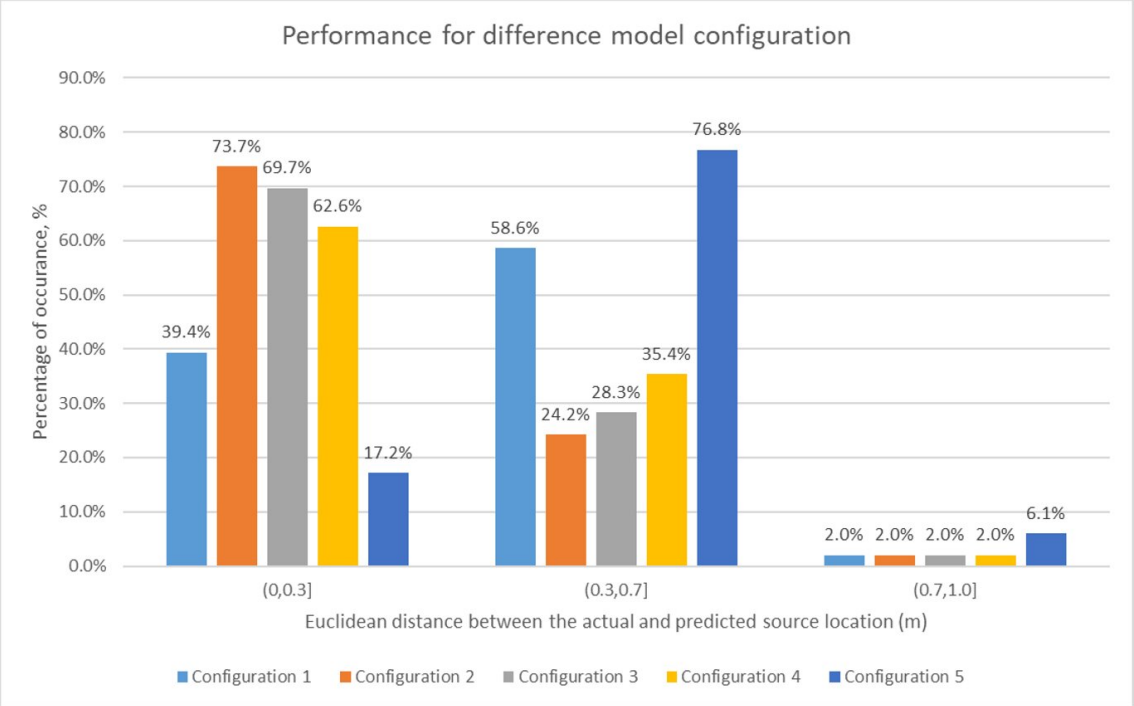


Figure 9: Comparison of model performance between the different model configurations

3.2 Model validation test

The data has been selected randomly from the validation dataset and fed into the model to observe its ability to predict the location of the ethanol vapor source in a small simulated office room. (Table 3) below shows the result of the model prediction in the terms of Euclidean distance error. To get a clear picture of the distance between the predicted and true gas emission location (Figure 11) illustrates the plotting of the predicted and actual location obtained from the table using matplotlib. As can be seen, six number of validation samples at different locations of the ethanol source have been fed into the model to predict the location of the gas source. From the table, most of the error of the y-coordinate output is larger compared to the x-coordinate output. The minimum and maximum Euclidian distance error between the actual and predicted gas source from the validation dataset is 0.20m and 0.34m respectively. The error presence might be due to the concentration of ethanol vapor for the certain location being similar to each other. Hence, within the range of the observed error, it is adequate for the gas localization task and it will make the task become much easier. It had proven that the DNN model able to predict the location of the gas source inside a simulated indoor environment.

3.3 Comparison with the ANN model

In this subsection, an ANN model with one hidden layer was trained using the same synthetic datasets. This is to observe the performance comparison between both models in the task of predicting the gas source location. (Figure 12) shows the performance comparison between the ANN and DNN model in terms of the percentage of occurrence for the model to predict the gas location. As can see

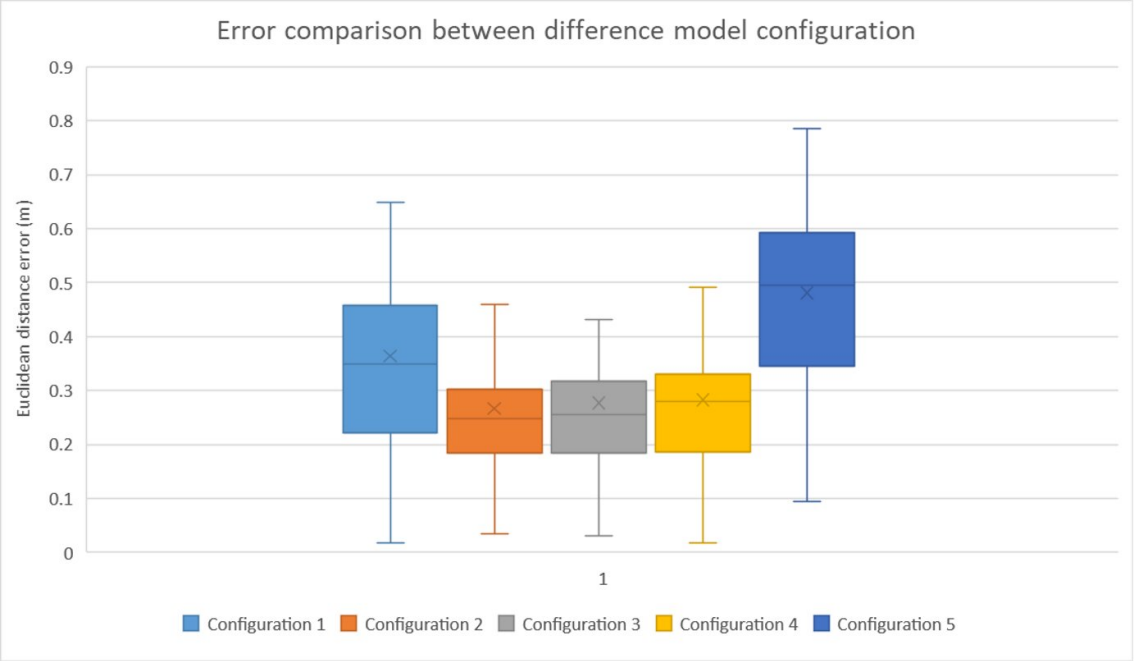


Figure 10: Box plot of Euclidian distance error for the different DNN model configurations

Table 3: Comparison between the actual gas source location and predicted gas source location

Coordinate of Actual Gas Source Location (m)		Coordinate of Predicted Gas Source Location (m)		X-error (m)	Y-error (m)	Euclidean distance (m)
x	y	$\hat{x}$	$\hat{y}$			
2.40	3.60	2.27	3.75	0.13	0.15	0.20
5.10	1.20	4.99	0.95	0.11	0.25	0.27
0.90	3.60	0.72	3.89	0.18	0.29	0.34
1.80	1.80	1.67	1.96	0.13	0.16	0.21
3.60	4.20	3.75	4.33	0.15	0.13	0.20
1.50	3.30	1.42	3.52	0.08	0.22	0.23

from the histogram chart, the DNN model is able to predict the gas source with Euclidean distance error in the range of 0.0m to 0.3m with 73.7% which is higher compared to the ANN model with only 52.9%. The ANN model predicts the location of gas within the range of 0.3m to 0.7m from the actual gas source with 46.2% of the total dataset. This result revealed that the DNN model has a better performance and accuracy compared to the ANN model in performing the gas source localization.

4 Conclusion

In this study, a DNN model was utilized to predict the location of the gas source inside a small simulated empty room after a sufficient training process using the TensorFlow package in the Google Collaboratory platform. During the training process of the model, it is highly dependent on the collected dataset. Therefore, CFD modelling was utilized to gain a sufficient dataset for the model. 361 number of simulated ethanol vapor dispersion at different locations were divided into the training

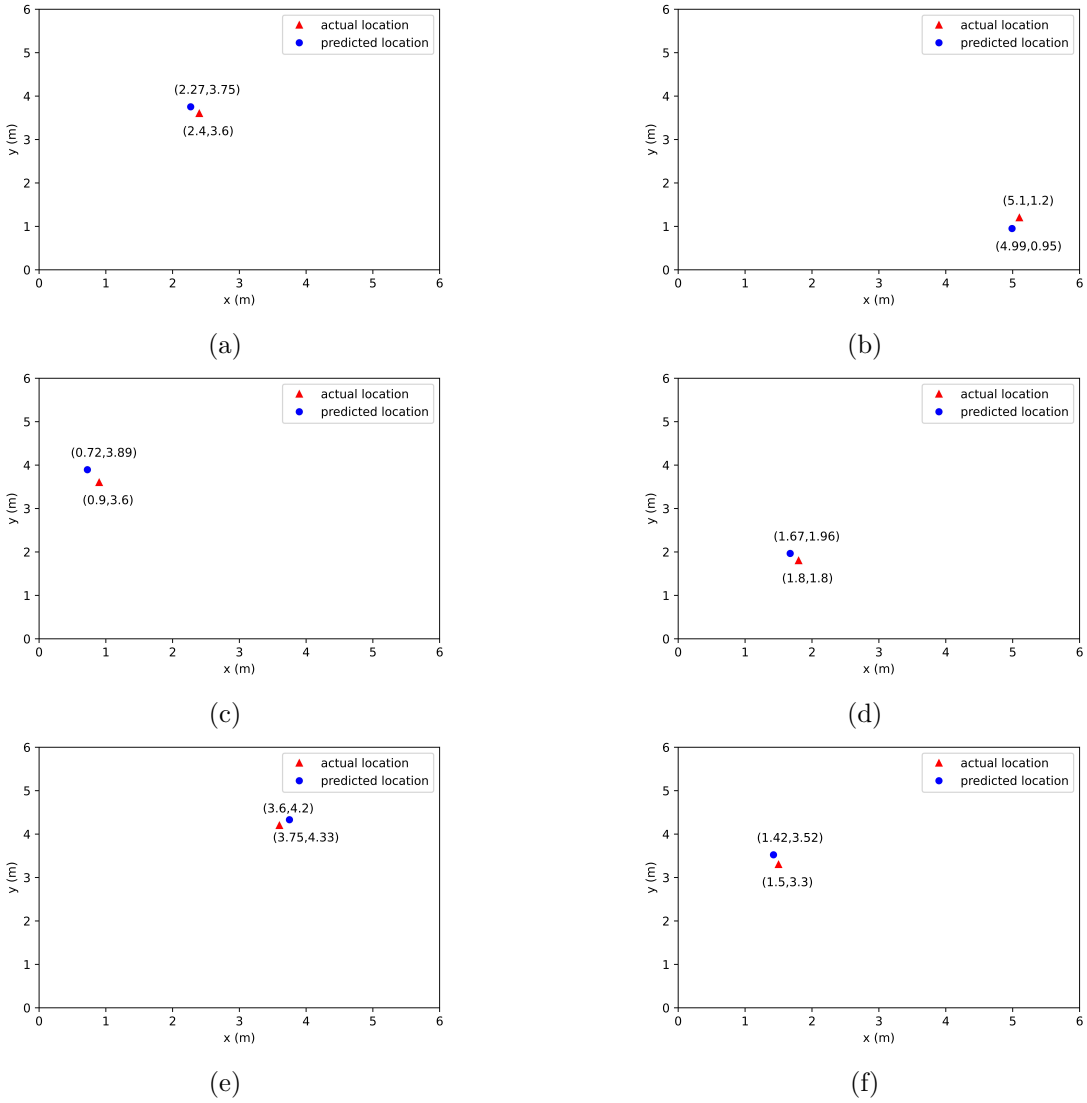


Figure 11: The plotting of the predicted gas source and actual gas source location from the validation dataset

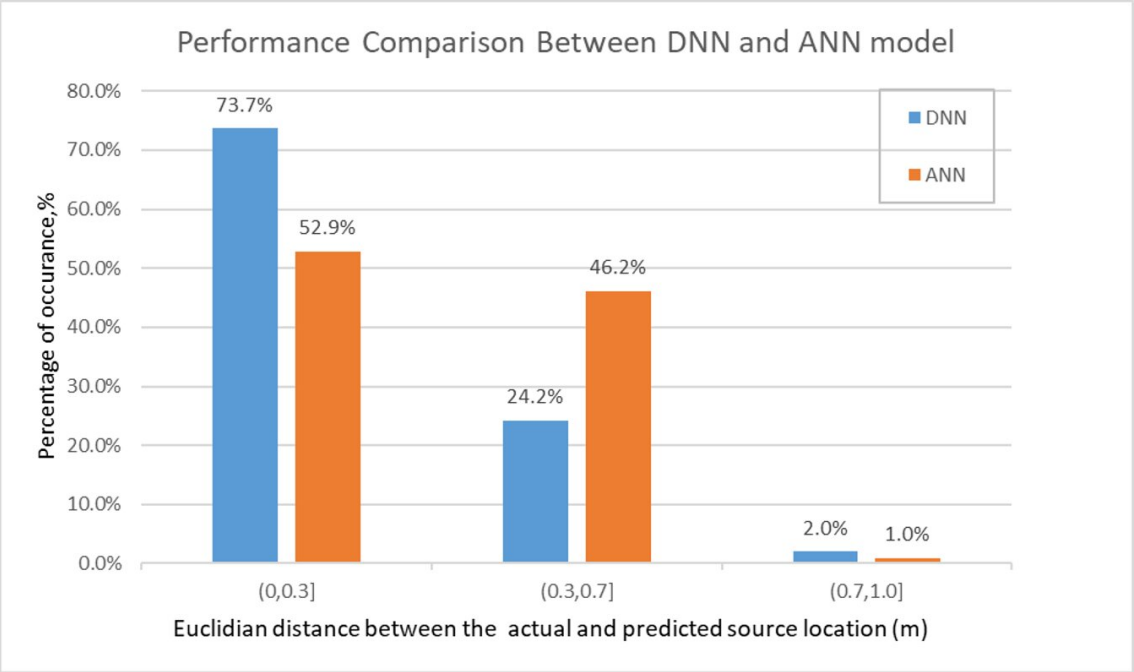


Figure 12: Performance comparison between DNN and ANN model

and testing datasets. The model was able to identify the location of the ethanol vapor emission source after the training process. This study also focused on the hyperparameter of the DNN model which is the number of nodes in the hidden layer. Several combination numbers of hidden layer parameters have been tested and it is found that the best configuration of the hidden layer for the first and second hidden layer of the model to predict the gas source in this study was 500 nodes and 300 nodes respectively. Therefore, the DNN model contains 19787 input nodes, 2 hidden layers, and 2 output neurons. The model has the minimum and maximum error in terms of Euclidean distance error within the range between 0.03m to 0.460m which is adequate to perform the gas source localization task. For average, the model has a mean error of 0.24m. In conclusion, the CFD-DNN method to localize the gas source inside an empty room shows low error and high accuracy.

In the future, a real gas sensor array will be used to collect the gas emission dataset in an indoor environment. This dataset will allow researchers around the world to train the deep learning model to obtain a gas localizing system that has been integrated with artificial intelligence for helping human beings to make a decision.

Funding

The author would like to acknowledge the support from the Graduate Research Assistant (GRA) Grant (9001-00682) from Universiti Malaysia Perlis (UniMAP). This research was also funded by the Malaysian Technical University Network (MTUN) Research Grant from the Ministry of Higher Education of Malaysia (MOHE) (9002-00094/9028-00002).

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

References

[1] G. Kowadlo and R. A. Russell. (2008). Robot odor localization: A taxonomy and survey, *The International Journal of Robotics Research*, vol. 27, no. 8, pp. 869–894, Aug. 2008, doi: 10.1177/0278364908095118

[2] T. I. Express.(2016). Bangladesh: 250 fall sick after toxic gas leak at fertiliser unit | World News,The Indian Express, https://indianexpress.com/article/world/world-news/bangladesh-toxic-gas-leak-at-fertiliser-unit-sickens-250-2992087/ (accessed Dec. 02, 2020).

[3] New Age.(2020). One in five fire accidents linked to gas leakage, https://www.newagebd.net/article/115736/one-in-five-fire-accidents-linked-to-gas-leakage (accessed Dec. 28, 2022).

[4] V. M. H. Bennetts, A. J. Lilienthal, A. A. Khaliq, V. P. Sese, and M. Trincavelli.(2013). Towards real-world gas distribution mapping and leak localization using a mobile robot with 3d and remote gas sensing capabilities, *Proceedings - IEEE International Conference on Robotics and Automation*, pp. 2335–2340, 2013, doi: 10.1109/ICRA.2013.6630893.

[5] V. H. Bennetts, M. Trincavelli, A. J. Lilienthal, V. Pomareda, and E. Schaffernicht.(2014). Online parameter selection for gas distribution mapping, *Sensor Letters*, vol. 12, no. 6–7, pp. 1147–1151, 2014, doi: 10.1166/sl.2014.3191.

[6] L. Dong, H. Zuo, L. Hu, B. Yang, L. Li, and L. Wu.(2017). Simulation of heavy gas dispersion in a large indoor space using CFD model, *Journal of Loss Prevention in the Process Industries*, vol. 46, pp. 1–12, 2017, doi: 10.1016/j.jlp.2017.01.012.

[7] H. Ishida, Y. Wada, and H. Matsukura.(2012). Chemical sensing in robotic applications: A review, *IEEE Sensors Journal*, vol. 12, no. 11, pp. 3163–3173, 2012, doi: 10.1109/JSEN.2012.2208740.

[8] H. Ishida, T. Yamanaka, N. Kushida, T. Nakamoto, and T. Moriizumi.(2000). Study of real-time visualization of gas/odor flow image using gas sensor array, *Sensors and Actuators, B: Chemical*, vol. 65, no. 1–3, pp. 14–16, Jun. 2000, doi: 10.1016/S0925-4005(99)00415-3.

[9] E. M. Moraud and D. Martinez.(2010). Effectiveness and robustness of robot infotaxis for searching in dilute conditions, *Frontiers in Neurorobotics*, vol. 4, no. MAR, pp. 1–8, 2010, doi: 10.3389/fnbot.2010.00001.

[10] J. Burgués, V. Hernández, A. J. Lilienthal, and S. Marco.(2020). Gas distribution mapping and source localization using a 3D grid of metal oxide semiconductor sensors, *Sensors and Actuators, B: Chemical*. , vol. 304, no. August, p. 127309, 2020, doi: 10.1016/j.snb.2019.127309.

[11] H. Ishida, K. Suetsugu, T. Nakamoto, and T. Moriizumi.(1994). Study of autonomous mobile sensing system for localization of odor source using gas sensors and anemometric sensors, *Sensors and Actuators A: Physical*, vol. 45, no. 2, pp. 153–157, 1994, doi: 10.1016/0924-4247(94)00829-9.

[12] J. Atema.(1996) Eddy chemotaxis and odor landscapes: Exploration of nature with animal sensors, *Biological Bulletin*, vol. 191, no. 1, pp. 129–138, 1996, doi: 10.2307/1543074.

[13] P. Neumann, V. H. Bennetts, and M. Bartholmai.(2012). Adaptive gas source localization strategies and gas distribution mapping using a gas-sensitive micro-drone, *Fachtagung Sensoren und Messsyst*. 2012, pp. 800–809, 2012, doi: 10.5162/sensoren2012/P5.4.

[14] M. Vergassola, E. Villermaux, and B. I. Shraiman.(2007). Infotaxis' as a strategy for searching without gradients, *Nature*, vol. 445, no. 7126, pp. 406–409, 2007, doi: 10.1038/nature05464.

[15] J. G. Monroy, J. L. Blanco, and J. Gonzalez-Jimenez.(2015). Time-variant gas distribution mapping with obstacle information, *Autonomous Robots*, vol. 40, no. 1, pp. 1–16, 2015, doi: 10.1007/s10514-015-9437-0.

[16] M. Reggente and A. J. Lilienthal.(2009). Three-dimensional statistical gas distribution mapping in an uncontrolled indoor environment, *AIP Conference Proceedings*, vol. 1137, pp. 109–112, 2009, doi: 10.1063/1.3156484.

[17] M. S. Awadalla, T.F. Lu, Z. F. Tian, and B. Dally.(2012). CFD modeling of 3D indoor gas contaminant plumes for testing search algorithms of mobile robot, *Gas*, vol. 2, no. S5, p. S6, 2012, [Online]. Available: http://www.cfd.com.au/cfd_conf12/PDFs/201AWA.pdf

[18] P. Ojeda, J. Monroy, and J. Gonzalez-Jimenez.(2021). Information-driven gas source localization exploiting gas and wind local measurements for autonomous mobile Robots, *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1320–1326, 2021, doi: 10.1109/LRA.2021.3057290.

[19] M. G. W Khalaf, C Pace.(2008). Gas detection via machine learning, *International Journal Computer Electrrical Autonomous Control Inf.*, vol. 2, no. 1, pp. 61–65, 2008.

[20] S. Mahfouz, F. Mourad-chehade, P. Honeine, J. Farah, and H. Snoussi.(2016). Machine learning in WSNs, vol. 16, no. 14, pp. 5795–5804, 2016.

[21] 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, *Computers and Chemical Engineering*, vol. 125, pp. 476–489, 2019, doi: 10.1016/j.compchemeng.2019.03.012.

[22] 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), vol. 18, no. 12, 2018, doi: 10.3390/s18124484.

[23] 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, *Digital Chemical Engineering*, vol. 3, no. March, p. 100029, 2022, doi: 10.1016/j.dche.2022.100029.

[24] Z. H. M. Juffry, K. Kamarudin, A. H. Adom, H. Nishizaki, A. Zakaria, S.M. Mamduh, A. N. Abdullah(2022). Deep neural network for localizing gas source based on gas distribution map, *International Conference on Electrical, Control and Computer Engineering*, 2022, pp. 1105–1115. doi: 10.1007/978-981-16-8690-0_96.

[25] P. K. Kundu, I. M. Cohen, and D. R. Dowling.(2016). Conservation laws, *Fluid Mechanics*, pp. 109–193. doi: 10.1016/b978-0-12-405935-1.00004-6.

[26] ANSYS.(2011). ANSYS CFD-Solver theory guide. Release 14.0, vol. 15317, no. November, pp. 724–746, 2011, [Online]. Available: http://www1.ansys.com/customer/content/documentation/140/cfx_thry.pdf

[27] J. C. R. Hunt.(1973). Mathematical models of turbulence, *Academic Press*, 169 pp. £2.50 or 7.50," J. Fluid Mech., vol. 57, no. 4, pp. 826–828, Mar. 1973, doi: 10.1017/S0022112073222048.

Journal's webpage: http://univagora.ro/jour/index.php/ijccc/

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Juffry, Z.H.M.; Kamarudin, K.; Adom, A.H.; Miskon, M.F.; Kamarudin, L.M.; Zakaria, A.; Mamduh, S.M.; Abdullah, A.N.(2023). Application of Deep Neural Network for Gas Source Localization in an Indoor Environment, *International Journal of Computers Communications & Control*, 18(3), 5084, 2023.

https://doi.org/10.15837/ijccc.2023.3.5084