Prediction of Volatile Organic Compounds (VOCs) From Decomposition of Local Household Food Waste Using the Artificial Neural Network

Siti Rohana Mohd Yatim, Ku Halim Ku Hamid, Kamariah Noor Ismail, Zulkifli Abdul Rashid, Nur Ain Mohd Zainuddin

Abstract: This study examines the potential of artificial neural network (ANN) to predict Total Volatile Organic Compounds (TVOCs) released via decomposition of local food wastes. To mimic the decomposition process, a bioreactor was designed to stimulate the food waste storage condition. The food waste was modeled based on the waste composition from a residential area. A feed forward multilayer back propagation (Levenberg – Marquardt training algorithm) was then developed to predict the TVOCs. The findings indicate that a two-layer artificial neuron network (ANN) with six input variables and these include (outside and inside temperature, pH, moisture content, oxygen level, relative humidity) with a total of eighty eight (88) data are used for the modeling purpose. The network with the highest regression coefficient (R) is 0.9967 and the lowest Mean Square Error (MSE) is 0.00012 (nearest to the value of zero) has been selected as the Optimum ANN model. The findings of this study suggest the most suitable ANN model that befits the research objective is ANN model with one (1) hidden layer with fifteen (15) hidden neurons. Additionally, it is critical to note that the results from the experiment and predicted model are in good agreement.

Keywords: Multilayer back propagation; waste storage, Volatile organic compounds (VOCs), local household food waste.

I. INTRODUCTION

Artificial neural network is an established branch of science, and is being dynamically developed for application in various practical fields. The networks constitute universal approximation systems that represent multidimensional data sets and these properties are often required in many practical applications in science. They are easily adapted to the changing environmental conditions.

Apart from its broad applicability, “Artificial Neural Network” (ANN) (from hereinafter should be referred as ‘ANN’) is able to generalize the acquired knowledge being; which, in this respect, is also known as a system of artificial intelligence ANN functions based on teaching algorithms that enable ANN to design appropriate topology of the network as well as to select parameters of the structure that are adjusted to the problem that needed to be solved [1]. It is acknowledged that ANN has been utilized in broad ways and these include; (1) estimation of global warming from temporary storage of municipal waste [2], (2) dioxin emission from incinerator [3], (3) modelling methane yield from organic fraction of municipal waste[4] and (4) modeling biological treatment processes in anaerobic condition [5]. ANN’s capability to simplify and learn new data is advantageous, as well as its ability to process nonlinear data and being highly tolerant to failures. Not only that, ANN can also be implemented with more than one layer for prediction purposes, for example the implementation of a three-layered ANN using pH and residence time as input for prediction of pH, acetic acid and propionic acid concentration in an acidogensi Apart from its broad applicability, c reactor [6].

Another example is the use of ANN’s feed forward back propagation to control the production of methane from a continuously stirred anaerobic tank reactor, which was operated with different organic loading rates [7]. The ANN model demonstrated ANN’s capability to effectively predict gas production and composition from the reactors. Ozkaya [8], on the other hand, presented ANN model for predicting the methane fraction in landfill where seven variables such as pH, alkalinity, Chemical Oxygen Demand, sulfate, conductivity, chloride and waste temperature were selected as inputs with methane gas as output. In their study, the ANN model was proven to be in a good agreement between experiment and prediction.

Another example is the use of ANN-back propagation method to project modeling for a temporary storage bins. In waste management, temporary storage bins could be a source of biogas emission and global warming. This study by Velumani [2] found that ANN-back propagation network is able to project models that helped decision makers select environment friendly options in the design of temporary storage to manage municipal solid waste in the local authority territory.
The ANN model demonstrated perfectly the use of ANN in accurately predicting the number of bins and global warming production from municipal solid waste generation rate. Based on these studies, it is clear that ANN has wide-ranging applicability due to its flexibility and capability in the discipline of science and technology. Therefore, ANN makes for a perfect tool for research aiming to model “total volatile organic compound” (TVOC) produced from the decomposition process of food wastes. In the subsequent section, the prediction of the models under dynamic circumstances will be presented and discussed.

II. MATERIALS AND METHODS

A. Bioreactor set up

This study was carried out to simulate the process that occurs during waste storage. Anaerobic decomposition food waste in a bioreactor (Fig. 1) was the chosen decomposition concept applied in this study, with major components that include; thermocouple, three gas collection points located at the top, middle and bottom. Meanwhile, the upper cover was secured, ensuring airtight conditions that prohibit infiltration of the surrounding air. The humidifier (MM Friocell, US) was used to incubate the bioreactor according to temperature set up. Two temperature sensors (K thermocouples, TMC6-HA) were placed at the center and the top of the FW. The food waste (FW) model was created based on the food waste (FW) composition in Puncak Alam’s household. The volume of FW model was made in quantities that are sufficient for experiment purposes. Based on the design of this study, approximately 15 kg of FW were placed in the bioreactor for up to 21 days at four different temperatures i.e. 20°C, 30°C, 40°C and 50°C. At each sampling point, volatile organic compounds (VOCs) samplings were carried out using head space technique. This involved the use of air sampling pumps (model LFS – 113c) that draws air into a carbo trap. The pumps were calibrated using DC-Lite Calibrator, model 717-01 (Skk Inc) before usage. The flow of each tube was adjusted at 50 ml/min. The suction point at the air sampler was connected to the sampling point at the bioreactor and another outlet point was connected to the flask (see Fig. 1).

![Fig. 1: schematic diagram of bioreactor](image)

B. Analytical

The analysis of VOCs was performed using Gas Chromatography Mass Spectrometry (GCMS), model Agilent Technologies, 7890A equipped with 5975G inert MSD with triple axis detector a 7683B series automatic injector made from USA. The GCMS was equipped with a fused silica capillary nonpolar column, Elite 5MS with 30 m length, 0.25 mm diameter and 0.25 µm film thickness (J & W scientific, USA). Helium was used as the carrier gas, with flow rate setting at 1.7 ml/min. Injection mode was set at split ratio of 1:15 and 1 µl for every samples made and later to be analyzed. The operating conditions for the oven were as followed: (1) the injector temperature and detector temperature were programmed at 240°C and 280°C, the column temperature was heated at 50°C and held for 5 minutes, then ramped at 5 °C/min to 280°C [9]. The total time needed for GC program to be completed was approximately 45 minutes. To perform control over GC/MS a program name “Chemstation” was chosen. An identification system (in the computer program) i.e. Automated Mass Spectral Deconvolution was opted to extract spectra for individual components in a GC-MS data file and identifies target compounds by matching these spectra against library [10].

C. Artificial Neural Network background and modelling

i. Software

The neural network model for decomposition process of FW was developed using Neural Network Toolbox software (version 7.9.0. (R2009b)). For this study, two software have been performed i.e. a feed forward neural network and the back - propagation learning algorithm ANN to develop ANN model for these two are the most commonly applied ANN layout [11].

ii. Preparation of Data

ANN modeling requires stringent data preparation, which was derived from the experiment results. The data were carefully selected as they may potentially affect the ability and reliability of the performance of neural network. The data required for running the neural network models include input data and output (or target) data. In this study, the parameters (input) selected include temperature, time, pH, moisture content, relative humidity and oxygen level. Meanwhile, the output parameter is total Volatile Organic Compounds (TVOCs). The input and target data were prepared in a Microsoft Excel Spreadsheet. The data set of eighty eight (88) samples were separated into three subsets (1) training set, (2) validation set and (3) testing set by random selection

iii. Architecture design

This step explains the construction of the neural network structure and its operation. Each ANN models are differently developed and for this study, the exact number of layers and number of neurons in each layer must be identified through “trial and error” method. In this study, the multilayered feed forward architecture was used for designing the artificial neural network models. The networks were designed to contain three (3) different layers namely; input layer, hidden layer and an output layer. From the results, it was noted that the input layer of the neural network models contained seven (7) neurons while the output layer consisted of one neuron. As for identification of the hidden neurons, a “trial and error” methodology is commonly used to find the optimal number of the hidden neurons.
Fig. 2 shows ANN schematic diagram that was used for prediction of this case study model. Each ‘X’ components represent different parameters and there were seven (7) parameters involved (X1 until X7) including time, outside temperature, inside temperature, pH, relative humidity, oxygen level and moisture contents. Meanwhile, “D” (refer to Fig. 2) represents the target output, which, in this case is total volatile organic compound (TVOCs) (ppm) collected from the experimental work. Fragment “E” (refer to Fig. 2) is a schematic diagram that represents ANN model with one hidden layer and several hidden layers and another one output layer with one neuron. Fragment “O” (see Fig. 2), on the other hand, represents the predicted TVOCs (ppm) computed from network. While weight connection in hidden layer is represented by fragment “W_h” (see Fig. 2) they represent, “W_o” represents weight connection in output layer, with fragment “b” representing bias. Fragment (f(ha)) indicates transfer function for each neuron in hidden layer, while tansig and fragment f(y) is transfer function in output layer, where pure linear was selected. In this study, the activation functions assigned in hidden layer of the neurons are called sigmoid functions (TANSIG) and the linear activation function is called (PURELIN) as output layer of the neurons as these two functions are commonly used [12].

**Fig. 2:** ANN schematic diagram used in this modelling iv. Neural network training

A set of input data with a total of eighty eight (88) samples were used as the input matrix (presented as: 88 X 7) and the corresponding values of Total Volatile Organic Compound (TVOCs) obtained from the experiment as target matrix (presented as: 88 X 1) in the developed ANN model. In this study, the architecture of the neural network models were determined through application of twelve (12) different number of hidden neurons in the hidden layers neurons [from seven (7) to eighteen (18) neurons] for three (3) different trials of data separation (70% of training set: 15% of validation set and the remaining 15% represents testing set). When the model no longer improves despite the increase of the hidden neurons, the model with the smallest amount and maximum performance were then chosen as the best model [13]. Performance and validation of the neural network must be achieved to determine the ANN’s ability to solve the problem. Hence, in this study, Mean Squared Error (MSE) and Regression (R) value are calculated to evaluate and validate performance of the neural network models using the following equation:

\[
x_{\text{norm}} = \frac{(X_{\text{real}} - X_{\text{min}})}{(X_{\text{max}} - X_{\text{min}})}
\]  

(\text{Eq1})

Where:

- \(X_{\text{norm}}\) : is a normalized value
- \(X_{\text{real}}\) : is a real value
- \(X_{\text{min}}\) : the minimum values for the variable X
- \(X_{\text{max}}\) : the maximum values for the variable X

### III. RESULT AND DISCUSSION

#### A. VOCs emission

Based from the results, the VOCs production in all samples started without delay time and reached the maximum emission within day 4-7 (see Fig. 3). The daily rate of VOCs produced at temperature 20°C, 30°C, 40°C and 50°C is presented in Fig. 3, which indicates that the highest production rate occurred at temperature 50°C and the least at temperature 20°C. VOCs production increased due to the microbial activities; which are also influenced by temperature. The decrease of VOCs after day 7 indicates possibility of thermal shock, as well as the decline of microbial population [14].

**Fig. 3:** TVOCs production at different temperature

#### B. Performance of developed ANN

In this study, the proposed ANN was feed forward Levenberg–Marquardt back-propagation algorithm, with tan sigmoidal as a transfer function and the number of hidden layer was selected as one. Details of the MSE from the training data set versus the selected number of hidden neuron units in hidden layer are presented in Table I. Based on Table I, fifteen (15) hidden neuron units were chosen for the training data set based on the minimum MSE value of 0.00012. According to Hu [15], if too many, hidden neurons will lead to over fitting.
Meanwhile, if only few hidden neurons are applied, this will then cause the ANN model to be unsatisfactorily flexible to demonstrate the experimental signal [16]. Besides that, Hu [15] also indicated that the determination of optimal amount needed for training data is challenging. Throughout the training and testing process, the training error decreases until over fitting phenomena occurs and that is when the error starts to increase [17]. In this study, to prevent over fitting, the training step was halted when the error value of the validation data set increased.

Table I: MSE Value from Training Data Set for Various Number of Hidden Neuron

<table>
<thead>
<tr>
<th>No of Hidden Neuron</th>
<th>Model Error (MSE Value)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0.00317</td>
</tr>
<tr>
<td>8</td>
<td>0.00383</td>
</tr>
<tr>
<td>9</td>
<td>0.00404</td>
</tr>
<tr>
<td>10</td>
<td>0.00402</td>
</tr>
<tr>
<td>11</td>
<td>0.00213</td>
</tr>
<tr>
<td>12</td>
<td>0.00035</td>
</tr>
<tr>
<td>13</td>
<td>0.00312</td>
</tr>
<tr>
<td>14</td>
<td>0.00276</td>
</tr>
<tr>
<td>15</td>
<td>0.00012</td>
</tr>
<tr>
<td>16</td>
<td>0.00587</td>
</tr>
<tr>
<td>17</td>
<td>0.00552</td>
</tr>
<tr>
<td>18</td>
<td>0.00625</td>
</tr>
</tbody>
</table>

Among all the variables dedicated for artificial neural networks (ANN), a feed forward back propagation of several variables ANN types, a feed forward back propagation multilayer network with Levenberg-Marquardt (LM) training algorithm; produced “minimal MSE” and “maximal correlation coefficient R” for the data set. It was noted that there were seven (7) numbers of the operating neurons in the hidden layer (see highlighted area in Table II) which were determined through minimization of MSE function. Table II shows characteristic of the developed ANN Model in this case study. In this study, the minimum value of MSE and the maximum value of R were considered as the best neural network model. Therefore, the structure of the utilized neural network had been configured with 7 input neurons, 7 hidden neurons in 1 hidden layer and one output neuron.

Table II: Characteristic of the Developed ANN Model

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Commentary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm</td>
<td>Feed forward back propagation</td>
</tr>
<tr>
<td>Minimized error function</td>
<td>MSE</td>
</tr>
<tr>
<td>Training algorithm</td>
<td>Levenberg-Marquardt</td>
</tr>
<tr>
<td>Learning type</td>
<td>Supervised</td>
</tr>
<tr>
<td>Hidden layer</td>
<td>Hyperbolic tangent transfer function</td>
</tr>
<tr>
<td>Output layer</td>
<td>Pure linear transfer function</td>
</tr>
<tr>
<td>Number of neuron in input layer</td>
<td>7</td>
</tr>
<tr>
<td>Number of hidden layer</td>
<td>1</td>
</tr>
<tr>
<td>Number of neuron in hidden layer</td>
<td>7</td>
</tr>
<tr>
<td>Number of neuron in output layer</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 4 illustrates the scatter regression (R) plot that represents the actual data from the experiment as well as the predicted TVOCs from the developed ANN model of the training data set. The solid linear line indicates the acceptable value between the network output (TVOCs from experimental work) and the target output (TVOCs of ANN model). Fig 4 also highlights best correlation coefficient between the ANN models predicted value and the experimental value, which is recorded at 0.9676. From the experiment results, a good correlation for the prediction is MSE with the value of 0.00012. This is not the first experiment involving ANN model for waste study. Another research by Boniecki [18] have used ANN model for the prediction of ammonia released from sewage sludge composting and have found MSE value as 0.0155 and R value as 0.9996 respectively. Another study, conducted by Akkaya [19], found that the correlation coefficient value of the training set is 0.991 for the prediction of heating value from solid waste. As mentioned by Shahanzadeh [20], the closer the R value to 1, the better the model fits to the experimental or actual data. Additionally, Sargolzaei [21] determined three conditions of ANN model based on R value; (1) if the R value more than 0.9 indicates a very satisfactory model performance, (2) if R value in the range 0.8-0.9 shows a good performance and (3) if R value less than 0.8 indicates insufficient model performance.
The performance of the optimal ANN was then certified by using validation data test with another new data set consists of 13 data points; one which have not been used during the training process. Fig. 7 shows a scatter plot for actual value and predicted value for VOCs from the validation data set. After analyzing the data using ANN, the $R$ value for the validation data set is 0.97267. This value demonstrated that the model has satisfactory predictive ability and of good quality.

Meanwhile, Fig. 8 presents graph for testing fitness of the developed ANN model. The testing data set shows, the developed ANN model is acceptable since the $R$ value is 0.99244. Fig. 9 shows a comparison of the volatile organic compounds both experimental and predicted values from developed ANN model for testing data set. The Fig. 9 represents the correlation coefficient with the experimental results where the regression $R^2$ value is 0.9815.

Meanwhile, Fig. 10 shows scatter regression plot of both the actual and ANN model predicted total volatile organic compounds value for all eighty eight (88) data points used in this study. The $R$ value found in this model is 0.99004, which indicates a very satisfactory model performance.

### Table III: The Mean Square Error (MSE) and Regression (R) Values for the Training, Validation and Testing

<table>
<thead>
<tr>
<th></th>
<th>MSE Value</th>
<th>R-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>0.00210</td>
<td>0.99676</td>
</tr>
<tr>
<td>Validation</td>
<td>0.01948</td>
<td>0.97267</td>
</tr>
<tr>
<td>Test</td>
<td>0.01657</td>
<td>0.99244</td>
</tr>
</tbody>
</table>
IV. CONCLUSION

The present work described an ANN approach for predictions of numerous parameters involved in decomposition processes from local household food waste. The network with the highest regression coefficient (R) is 0.9967 and the lowest Mean Square Error (MSE) is 0.00012 has been selected as the Optimum ANN model and the suitable ANN model that befits the research objective is ANN model with one (1) hidden layer with fifteen (15) hidden neurons. This finding suggests that the ANN is a useful tool in predicting TVOCs production from decomposition process. TVOCs were projected very well and the developed neural network models are therefore suitable in predictive decomposition process. With the proven applicability of this tool, waste management workers or waste management agencies should be able to carry out risk estimation exercises and provide necessary precautions for safety purposes.

ACKNOWLEDGMENT

This work was supported by the LESTARI grant (600-IRMI/Dana KCM5/3/LESTARI (1152/2017)). The authors are grateful to the Faculty of Health Sciences and Faculty of Chemical Engineering for supporting this project. Authors do not possess any conflict of interest.

REFERENCES


AUTHORS PROFILE

Siti Rohana Mohd Yatim is a senior lecturer in Faculty of Health Sciences at Universiti Teknologi MARA, Puncak Alam Campus, Selangor. She obtained her Degree in Environmental Technology from UiTM and Master in Environmental Science from Universiti Putra Malaysia. She specialises in waste management, air quality and environmental health and safety. She has over 10 years’ experience in academia and teaches various environmental health and safety subject such as integrated solid waste management, wastewater management, vector control, environmental health ethics, pollution control etc. She is a member of Integrated methane research group (IMRGGE), which focuses on research on mosquito and other pest control; as well as a member of Malaysian Association of Environmental Health (MAEH).

Prof. Dr. Ku Halim Ku Hamid is a senior professor of Faculty of Chemical Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia. He received degrees of BSc and MSc from National University of Malaysia and PhD from The University of Sheffield, UK. His research interest are catalyst and catalysis, waste management, green energy, biotechnology, advanced material, natural product extraction, integrated biological treatment of wastewater. He has published about 250 technical papers, owner 26 of patents and had received more than 300 of invention and innovation awards from various parts of the world including award from World Intellectual Property Organization (WIPO) and outstanding inventor in Malaysian Technology Expo 2017. He also was appointed in various technical expert committees in various government agencies of Malaysia such as Ministry of Education, Ministry of Science, Technology and Innovation and Ministry of Urban Wellbeing, Housing and Local Authority. His relationship with industries makes him comprehend the needs of commercial elements and to inculcate in his R&D activities.

Published By: Blue Eyes Intelligence Engineering & Sciences Publication
Besides that, he has supervised more than 70 postgraduate students and being appointed as external examiners from various universities. His latest commercialized products are catalyst free technology of continuous biodiesel production and production technology of agarwood distilled water as health supplement.

**Associate Prof Dr Kamariah Noor Ismail** is currently a senior lecturer at Faculty of Chemical Engineering, Universiti Teknologi MARA (UiTM). She received her First Class BSc (Hons) in Chemistry from National University of Malaysia (UKM) in 1992. She then pursued her MSc (Eng) in 1993 at University of Leeds, UK under SIRIM sponsorship. She joined SIRIM as a Researcher in 1994 until 2001. Within this 7 years, she has gained a numerous of research and consultancy works. Associate Prof Dr kamariah joined UiTM at the end of 2001 and received her PhD in Chemical Engineering from the same university in 2009. Her main research interests are heterogeneous catalyst technology for toxic gas pollution abatement and palm oil craking, advanced materials development and wastewater treatment. As part of an academic contributions to the university and national, she was appointed as an external examiner of PETRONAS Technology Institute (2010-2012). She was also appointed as examiner for matriculation chemistry paper by Malaysian Education Ministry Department in year 2011. Currently she is a Head of Quality and actively involved in research supervision for undergraduates and postgraduate students. Since 2005 until now, she has granted numbers of research grants at amount of more than RM 800,000.00 She is now registered as associate member of Malaysian Institute of Chemistry (AMIC).

**Ir Dr. Zulkifli Abdul Rashid** P.Eng, C. Eng MIChemE is a senior lecturer at UiTM. He attained his first degree in chemical engineering at University of Manchester Institute of Science and Technology (UMIST) UK. After graduated, Ir Dr Zulkifli works as a project engineer and project manager for about 3 years in a few oil and gas companies. He then served as Environmental Control officer at DOE Malaysia. During his service at Department of Environment (DOE), he directly involved in various enforcement activities such as prosecutor officer, assess hazardous material installation projects, river rehabilitation project evaluation, analyze the reliability and performance of industrial wastewater and air pollution control system. Whilst at DOE, Ir Dr Zulkifli pursued his study in MSc Eng (Env Eng) at UPM and after 5 years, he joined UiTM. Ir Dr Zulkifli holds hsi PhD from UTP in Process Safety and Loss Prevention. In UiTM, Ir Dr Zulkifli involved in various projects such as assessment RDF technology for KPJT, research associated with process safety and quantitative of risk assessment, accident prevention and modelling, waste to energy, hazardous waste management, CMAH consultancy works, major hazard installation (MIH), collaborative research in industrial safety and environmental management at local and overseas. Ir Dr Zulkifli has published 2 books, 3 chapter in books, 10 technical report in MIH, and more than 50 research articles in journals and proceedings. Ir Dr Zulkifli is a Chartered Engineer member of MIChemE, UK (99919241), registered professional engineer (chemical engineer) to Board of Engineers Malaysia PE 19535 Safety and Health Officer (SHO) competence (SHO3114-J) and CMAH competent OKMH: HQ/16/MH/00/201).

**Nur Ain Zainuddin** received her B.Eng. Degree and M.Sc. in Chemical Engineering from Universiti Teknologi MARA, Malaysia. Currently, she is Lecturer at Faculty of Chemical Engineering, Universiti Teknologi MARA, Malaysia. Her general research interest in industrial wastewater treatment. N.A.Zainuddin is registered as graduate engineer under Board of Engineers Malaysia (BEM)