International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Development of Urine Alcohol Content Predicting System Using Machine Learning Based on the Electronic Nose

Hanis Amalia Saputri[1], Alexander Agung Santoso Gunawan*[2], Izzi Dzikri[3]

Submitted: 13/11/2022 **Accepted**: 15/02/2023

Abstract: Because many countries still rely on alcohol due to strong drinking cultural practices, efforts to minimize its detrimental usage are more difficult even though alcohol can affect public health. As a result, a device to track alcohol consumption is needed. This research developed an electronic nose based on Internet of Things (IoT) as a new potential tool for measuring the alcohol content in the human body from urine odor. The electronic nose prototype used three gas sensors and was tested using simulated urine that had nine predetermined alcohol contents. We developed several regression models for predicting urine alcohol content using machine learning algorithms, including linear and non-linear algorithms. Based on our experiments, we can reach satisfied results with Support Vector Regression (SVR) (MSE = 0.009) and Random Forest (MSE = 0.014). The results make the electronic nose prototype suitable for measuring the alcohol content in urine.

Keywords: *electronics nose, urine alcohol content, regression model machine learning*

1. Introduction

It was reported in the 2018 WHO global status report on alcohol and health that in 2016 there were 3 million deaths globally (5.3% of all deaths) caused by the harmful use of alcohol. Alcohol consumption had a more significant mortality effect than tuberculosis (2.3%), HIV/AIDS (1.8%), and diabetes (2.8%). The percentage of 5.3% of deaths caused by alcohol is likely to increase due to the customs and culture adopted in a country. Deaths from alcohol are not just a problem in one country. In some countries, such as Germany, which in 2012 was nominated as the population with the highest alcohol consumption. A total of 345,000 patients were treated for alcohol-related disorders. The low level of awareness of alcohol use reaches 80% of people with alcohol use problems who still have not received treatment for their harmful use or dependence [1]. Deaths caused by alcohol have both short-term and long-term consequences. Long-term alcohol use can increase chronic physical and mental health morbidity. These consequences include gastrointestinal disorders, cardiovascular disease, cancer, and other diseases [2]. Meanwhile, the short-term effects of alcohol can lead to a loss of self-control, which increases motorcycle accidents and violence and even causes alcohol poisoning. Therefore some countries, such as Australia, have established standard guidelines for alcohol consumption to reduce the risk of alcohol consequences in the short and long term [3]. However, there are studies that find no differences by sex, education, age, or ethnicity in the proportion of beliefs that alcohol is a risk factor for cancer [4]. Another study showed that low self-awareness was associated with short-term risk of harm by drinkers [5]. Furthermore, there are also studies that state that less than half of alcoholics are aware that alcohol causes cancer [4]. Reducing the short-term effects caused by alcohol can be done by checking the content in the body. There are several ways to detect alcohol, including urine, hair, breath, oral fluids, blood, and several other parameters. Recently, blood alcohol detection has become one of the most widely used due to well-established laboratory test methods, but detection by this method has a relatively high cost, and equipment availability is lacking in certain areas [6]. In addition to blood, the parameter that can be considered in dealing with short-term problems quickly is breathing. However, this parameter has drawbacks, such as the results of the tests carried out cannot be used as evidence in court [7]. Examination through urine can be a solution because the examination is relatively cheap and easy to obtain. In addition, urine can also be used as evidence in court when damage is obtained from alcohol users when they lose self-control. Therefore, this study will use urine as an experimental parameter and focus on the contribution of urine examination innovation by utilizing the Electronic Nose (e-Nose) for the alcohol examination process.

E-Nose is a device that tries to imitate the structure and function of the human nose that can be used to identify volatile substances or substances that quickly turn into vapor [8]. The way the E-nose works is to read the sensor response from the observed scent. Then the sensor response data will be analyzed for pattern recognition using a certain algorithm [9]. The implementation of e-Nose accuracy can be measured using a prediction algorithm. Regression algorithm prediction is required with several parameters in the model to optimize the prediction results because it only uses traditional regression algorithms. The resulting predictions are unsatisfactory and also require a long training time [10].

1 Computer Science Department, Binus Graduate Program, Master of Computer of Science, Bina Nusantara University, Jakarta, Indonesia, 11480
ORCID ID : 0000-0002-4068-9653
2,3 Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia, 11480
ORCID ID : 0000-0002-1097-5173
** Corresponding Author Email: aagung@binus.edu*

The use of an electronic nose (E-nose) with traditional regression prediction algorithms can produce unsatisfactory predictions and also takes a long time to train. Therefore, a regression prediction algorithm is needed with several parameters in the model to optimize the prediction results [10]. Several studies are experimenting with the device used to detect diabetes in the urine of patients. By leveraging multiple gas sensors and multiple machine learning algorithms, the device can detect simulated diabetic urine by adding variations of dextrose. The results of the classification of urine samples, K-Nearest Neighbors, have 100% accuracy, 100% Support Vector Machine, 96% Decision Tree, and 96% Random Forest [11]. Such as gas sensors used during the development and research process, namely urine containing alcohol for experiments with the equipment used. This experiment simulates synthetic urine and adds various alcohol content. Several models are used to develop to predict urine alcohol content accurately, including K-Nearest Neighbor (KNN), Random Forest (RF), Support Vector Regression (SVR), and Multiple Linear Regression (MLR). The main contribution is to carry out synthetic urine experiments, in which the alcohol composition in urine can be controlled and used as baseline truth data. This research is structured as follows; section 2 deals with the preparation of the data sets, section 3 deals with the research methods, section 4 deals with the presentation of the findings and discussion, and section 5 contains the conclusions.

2. Datasets Preparation

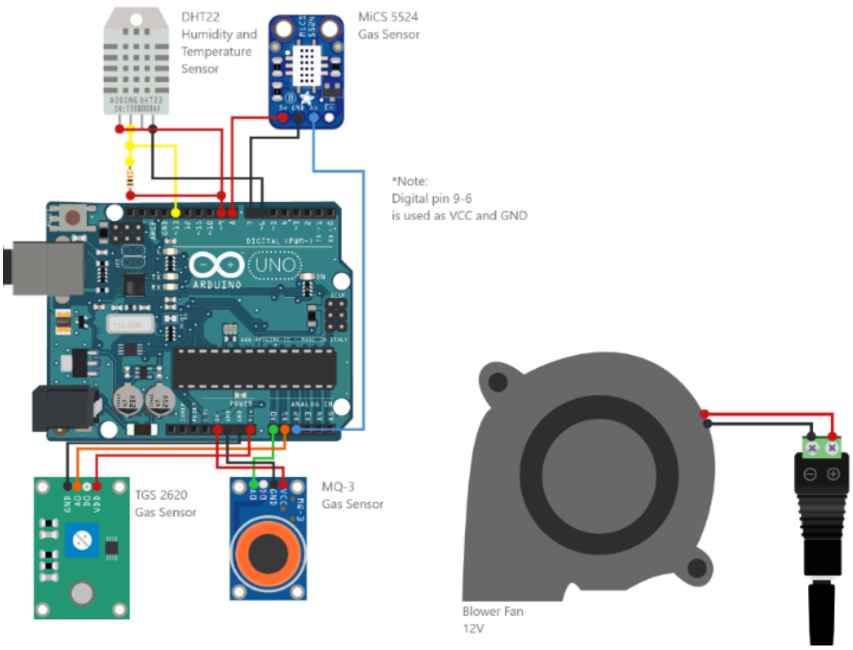
There are several stages in the preparation of the dataset, as follows.

2.1. Device Construction

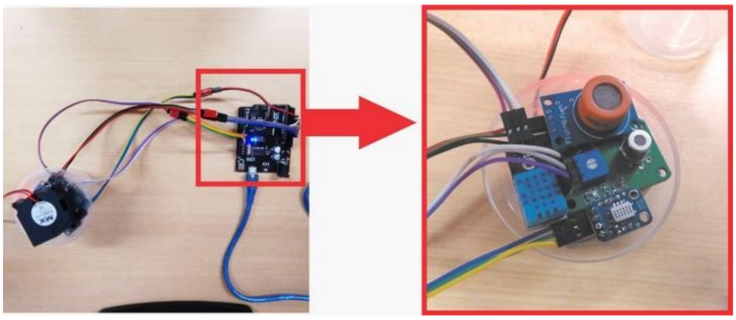
Three sensors were chosen because of their sensitivity to alcohol to make the device, including MQ-3, TGS-2620, and MiCS-5524. The three gas sensors are represented as S1, S2, and S3. DHT22 sensor (denoted by S4) measures the ambient temperature and humidity. A fan is also added to the housing to direct the airflow toward the detector. Figure 1 shows the finished device and the schematic design of the device.

2.2. Creating Sample

Considering the complex composition of human urine is difficult to imitate and the possibility of inconsistent values, synthetic urine is implemented based on the elemental composition of normal urine as an experimental sample. Drug testing laboratories have frequently used synthetic urine (SU) as a matrix to prepare quality control and other research [12]. Other studies have shown that the use of synthetic urine has been shown to accurately mimic real urine, even normal urine and sick human urine [13]. The composition of urine used in this study is shown in Table 1.



(a)



(b)

Fig 1. The prototype: schematic design (a) and hardware example (b)

Table 1. Simulated urine composition

Chemical	Concentration (%)
Water	95
Urea	2
Potassium	0.6
Chloride	0.6

The simulated urine diluted 70% ethanol into nine various concentrations (0.7%, 1%, 1.5%, 2%, 3%, 3.5%, 3.7%, 3.9%, and 4.1% v/v). Some levels are based on the urine alcohol content of a moderate drinker, and some are based on the alcohol content of the beverages [14]. These are the following formula to dilute the solution shown in Equation 1 and use only 50 ml of it for that collection.

$$V_1C_1 = V_2C_2 \tag{1}$$

The concentration and volume are denoted as C and V, respectively, while the subscription number 1 and 2 represent the solution state before and after being diluted, where the states are its concentration and volume. Using these various concentration samples is also proven effective in predicting by machine learning machines, that is also has been tested in the beverage sample such as beer [15].

2.3. Data Collection and Preprocessing

Processing software is used to collect the response of each sensor and save it into a CSV file. The pre-prepared urine synthetic solution was put into a 150 ml container with a gas sensor attached to the lid. The data collection was conducted for 10 minutes per second by taking nine times for each sample. Get approximately 10 min x 60 s x 9 samples or as many as 5400 total data points. Before data collection, each gas sensor was preheated for approximately 10 minutes until each sensor reached its optimum temperature before being operated. After each sensor was heated, put 50 ml of solution each into the container, then ventilated the odor until it stabilized the sensor response before changing it to the following sample.

2.4. Developing Model

Several machine learning algorithms were used to compare to find the most suitable model for measuring urine alcohol content. Regression methods allow the construction of linear and non-linear models to predict quantitative values that do not require a training base with a broad spectrum of labeled classes. The data collected from several sensors are used as a model feature or independent variable. In this study, four machine learning algorithms were used, including Multiple Linear Regression (MLR), Support Vector Regression (SVR), K-Nearest Neighbor (KNN), and Random Forest (RF).

2.5. *Tuning Hyperparameter*

Grid search cross-validation is implemented to select the best hyper-parameter for each model. *Cross-validation* is a re-sampling procedure used to test the machine learning model, while grid search is used for picking the best hyperparameter. The cross-validation procedure has a single parameter called K that divides the data into K groups, resulting in k-fold cross-validation. K-fold cross-validation is a common technique that divides a set of m examples into K equal-size sets (folds) of size m/K. The remaining sets are used to train models [6]. As for the grid, the search is also exhaustive for selecting a model. Since the approach will try every combination of the hyperparameter that has been provided. For each model, 5-fold for the cross-validation is the best hyper-parameter, which is shown in Table 2.

Table 2. The best hyperparameter for each model

Ml Model	Best Hyperparameter
MLR	None
SVR	Kernel = rbf, C = 10, gamma = 100
KNN	n neighbors = 3
RF	Max depth = 9, Mac feature = auto, n estimator = 100

3. Results and Discussion

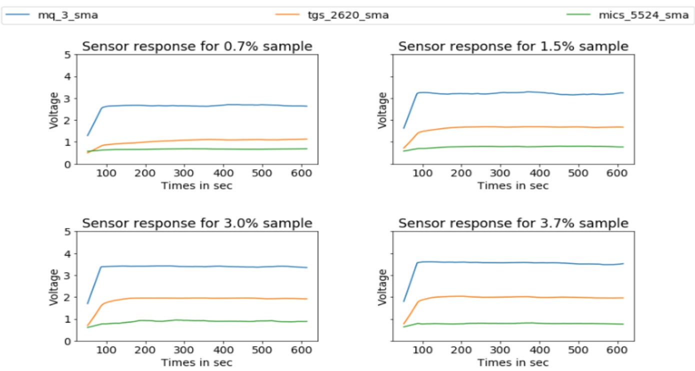


Fig 2. The Sensors response to each sample

The collected data contained noise resolved by implementing the simple moving average filter to smoothen the signal response. Filters are generally the simplest, fastest, and easiest to apply. This experiment used a window size of 50 to smooth the response data, and some graphs are shown in Figure 2.

After building and testing the machine learning model, the result shows that SVR outperforms the rest of the model. The following table summarizes the result of confidence and loss for each machine learning model.

Table 3. The confidence and loss value for each model

ML Model	R2 Score	MSE
MLR	0.689	0.485
SVR	0.993	0.009
KNN	0.981	0.028
RF	0.991	0.014

The train data curve shown in Figure 3 shows that none of the models are overfitting. Predictive results obtained to detect the value of alcohol in the urine do not correlate with the level of a person's drunkenness when consuming alcohol. However, looking at the cycle of the percentage of alcohol values in the body can play an essential role in seeing the correlation between a person's level of drunkenness [7]. The fluctuating data that may be generated from urine when alcohol is present in the system could be vital to understanding the alcohol cycle correlation in our system.

Author contributions

Hanis Amalia Saputri: Methodology, Verifying and Correcting Original Draft, Preparation, Implementation of Machine Learning Models.
Alexander Agung Santoso Gunawan: Methodology, Verifying and Correcting Original draft, Validate Machine Learning Results.
Izzi Dzikri: Validate Machine Learning Results, Writing Original Draft.

Conflicts of interest

The authors declare no conflicts of interest.

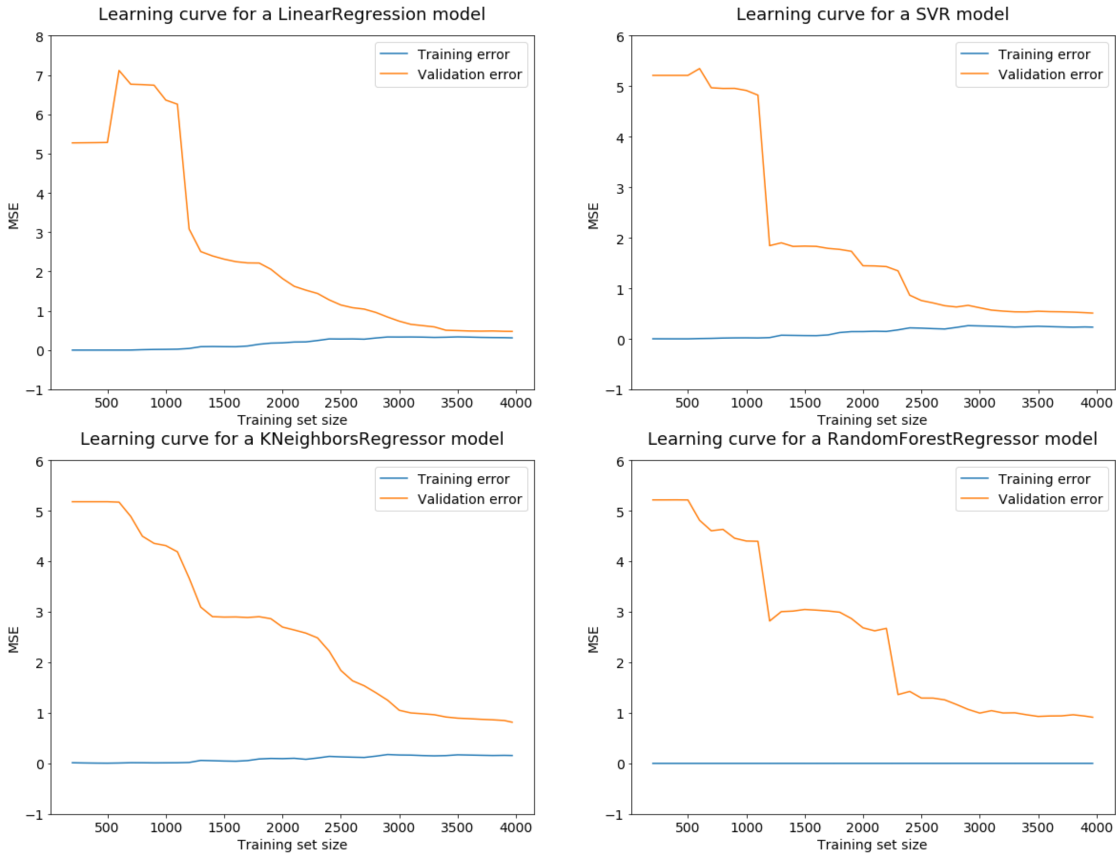


Fig 3. The learning curve for each model

4. Conclusion

Based on the research, it can be concluded that an electronic nose with multiple gas sensors was a reliable tool for measuring the alcohol content in a simulated urine sample. Each simulated model shows that the SVR outperforms the other models by producing an MSE of 0.009.
The tool newly created in this study can only predict the level of alcohol in the urine and cannot determine people's drunkenness level. In further research, this tool can be used to determine the correlation between the level of drunkenness in people and the percentage cycle of alcohol in the body.

References

[1] Batra A, Müller CA, Mann K, Heinz A. Alcohol Dependence and Harmful Use of Alcohol. Dtsch Arztebl Int 2016;113:301–10. https://doi.org/10.3238/arztebl.2016.0301.
[2] Madden M, Morris S, Stewart D, Atkin K, Gough B, McCambridge J. Conceptualising alcohol consumption in relation to long-term health conditions: Exploring risk in interviewee accounts of drinking and taking medications. PLoS One 2019;14:1–20. https://doi.org/10.1371/journal.pone.0224706.
[3] Coomber K, Mayshak R, Curtis A, Miller PG. Awareness and correlates of short-term and long-term consequences of alcohol use among Australian drinkers. Aust N

Z J Public Health 2017;41.
[4] Cotter T, Perez D, Dunlop S, Kite J, Gaskin C. Knowledge and beliefs about alcohol consumption, longer-term health risks, and the link with cancer in a sample of Australian adults. N S W Public Health Bull 2013;24:81–6. https://doi.org/10.1071/NB12089.
[5] Hasking P, Shortell C, Machalek M. University students' knowledge of alcoholic drinks and their perception of alcohol-related harm. J Drug Educ 2005;35:95–109. https://doi.org/10.2190/9Y34-F5XR-AQV5-KEL8.
[6] Andresen-Streichert H, Müller A, Glahn A, Skopp G, Sterneck M. Alcohol Biomarkers in Clinical and Forensic Contexts. Dtsch Arztebl Int 2018;115:309–15. https://doi.org/10.3238/arztebl.2018.0309.
[7] Hadland SE, Levy S. Objective Testing: Urine and Other Drug Tests. Child Adolesc Psychiatr Clin N Am 2016;25:549–65. https://doi.org/10.1016/j.chc.2016.02.005.
[8] Santos JP, Lozano J, Aleixandre M. Electronic Noses Applications in Beer Technology. Brew. Technol., InTech; 2017. https://doi.org/10.5772/intechopen.68822.
[9] Modak A, Roy RB, Tudu B, Bandyopadhyay R. A novel fuzzy based signal analysis technique in electronic nose and electronic tongue for black tea quality analysis. IEEE; 2016.
[10] Wang Y, Jia P, Cui H, Peng X. A Novel Regression Prediction Method for Electronic Nose Based on Broad Learning System. IEEE Sens J 2021;21:19374–81. https://doi.org/10.1109/JSEN.2021.3090449.
[11] Fadholi GUMELAR S, Mustafa BUDIYANTO H, Fawwaz MAYDA M, Alldino Ardi SUMBODO B. Exploration of Electronic-Nose Potential as Diabetes Urine Detection using Machine Learning Algorithms. Int J Adv Res Sci Eng Technol 2019;6.
[12] Kim VJ, Okano CK, Osborne CR, Frank DM, Meana CT, Castaneto MS. Can synthetic urine replace authentic urine to "beat" workplace drug testing? Drug Test Anal 2018:331–5. https://doi.org/10.1002/dta.2497.
[13] Shmaefsky B. Artificial Urine for Laboratory Testing-Revisired. JSTOR 2015;57:428–30.
[14] Jones AW. EXCRETION OF ALCOHOL IN URINE AND DIURESIS IN HEALTHY MEN IN RELATION TO THEIR AGE, THE DOSE ADMINISTERED AND THE TIME AFTER DRINKING. 1990.
[15] Jordan Voss HG, Mendes Júnior JJA, Farinelli ME, Stevan SL. A prototype to detect the alcohol content of beers based on an electronic nose. Sensors (Switzerland)2019;19. https://doi.org/10.3390/s19112646.