Detection Of Fluoride Content In Black Tea Using Electronic Nose

Istifa Shania Putri [a*], Danang Lelono[b]

Department of Computer Science and Electronics, FMIPA UGM, Yogyakarta, Indonesia
[a)] istifa.shania.putri@mail.ugm.ac.id
[b)]danang@ugm.ac.id

Abstract. The impact of fluoride on human health is still a controversial issue. The World Health Organization (WHO) has provided permissible levels of fluoride in black tea, but various commodities still ignore it. The purpose of this research is to design a software system used to build a predictive model for fluoride concentration based on the Rs / Ro ratio. The system uses input data in the form of e-nose sensor response and uses the Partial Least Square (PLSR) regression method for quick analysis of fluoride levels in various quality black tea (kw1, kw2, kw3). Identification of volatile compounds in black tea with various qualities was carried out using an e-nose equipped with 12 gas sensors. A prediction model for fluoride concentration based on the Rs / Ro ratio was successfully developed for each quality. The predictive performance values of fluoride concentration in the sample against the reference fluoride concentration Rs / Ro are: R2predkw1 = 0.986620 and RMSEpredkw1 = 0.004642; R2predkw2 = 0.994791 and RMSEpredkw2 = 0.010772; R2predkw3 = 0.994054 and RMSEpredkw3 = 0.005324. This shows that the indicators of predictive goodness for each black tea quality are satisfactory.

INTRODUCTION

The impact of fluoride on human health is still a controversial issue. Sufficient fluoride intake is recognized to have beneficial effects such as reducing dental caries in children. However, excess intake increases cases of dental and bone fluorosis. One source of fluoride intake is tea. Tea's fluoride content is found in a variety of compounds form [1]. Tea is a popular drink in the world community. Indonesia is one of countries who produced tea in large quantities. The Indonesian Central Bureau of Statistics in 2020 wrote that the volume of black tea exports reached 82.49% of total volume of tea exports. This shows that Indonesia's tea exports are mostly in the form of black tea. The safety standard for fluoride levels in black tea has been determined. The WHO Fluoride & Oral Health 2004 organization states that the fluoride level per brew of tea should not be more than 8.6 ppm [2]. Even so, the rules for fluoride levels in black tea are often ignored, so an instrument that can detect fluoride levels in black tea is needed. Analytical chemistry instrument, Ion-Selective Electrode (ISE) has been applied to identify fluoride content in black tea. The measurement outcomes offer a high level of accuracy but come at a greater cost in line with the instrument's requirements.

The Electronic Nose (e-nose) is an instrument that works like a human sense of smell by distinguishing samples based on smell, consisting of hardware systems and pattern recognition as artificial intelligence so that the instrument is able to recognize aroma and volatile compounds. E-nose offers the benefit of uncomplicated pre-processing, relatively rapid processing time, and cost-effectiveness. The main hardware components of the e-nose itself consist of a non-specific gas sensor array, which interacts with various types of chemicals with their respective sensitivity levels. Consequently, following this principle, the presence of other molecules can be minimized by leveraging the sensitivity of sensors that are responsive to compounds containing targeted element that to be measured. The main hardware component of the e-nose itself is an array of non-specific gas sensors, namely sensors that interact with various chemicals with their respective sensitivity levels, so based on this, the presence of other molecular factors can be

reduced by utilizing the sensitivity of the sensor, that sensitive to compounds containing elements to be measured. Apart from hardware, the e-nose has a pattern recognition system for analyzing the aroma of samples, such as Partial Least Square Regression (PLSR). The PLSR method is a robust multivariate analysis of multicollinearity. Meanwhile, multicollinearity is one of the main problems faced when analyzing gas sensor array data [3].

In black tea samples, e-nose has been used, including tea classification based on type [4]; black tea classification using ICA and SVM [5]; black tea classification based on planting location [6]. Based on this research, the scent pattern formed from detection using the e-nose is used qualitatively for identification or classification of samples. Because the aroma pattern is used qualitatively, predicting the concentration of fluoride in black tea requires a system capable of quantification. So that in this study, the researchers developed a system for detection of fluoride compounds in black tea using the e-nose combined with the PLSR analysis method.

RESEARCH METHODS

Analysis of the system

The fluoride present in tea exists not as an single element, but rather in a variety of compound forms [1], including Polycyclic Aromatic Hydrocarbons (PAHs) which can evaporate with other volatile compounds [7]. This leads to the tea's scent being composed of a variety of volatile compounds. So we need sensors that capable to capture the scent. The electronic nose (e-nose) is an instrument based on a gas sensor array that can capture the aroma formed by the volatile gases of the sample into patterned electrical quantities. Through the process of signal analysis with a pattern recognition machine, samples can be identified or classified. So far, the sensors used in the array have usually been general-purpose, unselected so that it is not focused on certain gas components. The advantage of using an unselected gas sensor allows the instrument to analyze various aroma characteristics of the sample. However, if the resulting aroma pattern is deconstructed to extract a gas with a particular concentration, it is then required to conduct further analysis using pattern recognition methods. The e-nose instrument utilized in this study is equipped with a array of twelve gas sensors, each sensor has its own function and characteristics. So that at the beginning of the measurement, each sensor provides a different initial value (baseline). Therefore we need a signal pre-processing method to equalize the initial value of each sensor. Furthermore, the signal from the baseline equalization needs to be normalized in order to minimize sensor deviation and differences in sensor scaling. Normalization that is carried out must make the data centralized so that the data meets the requirements as input into the regression analysis.

The sensor response has been equated to the initial value and has been centered, then a pattern recognition analysis method is needed to convert the sensor response into a target gas concentration. A pattern recognition method in the form of regression with robust analysis of multicollinearity, where multicollinearity is one of the main problems faced when analyzing gas sensor array data [3]. The existence of multicollinearity causes the estimated parameter to be bigger than it should be, thus the accuracy of the estimation will decrease. Meanwhile, multicollinearity itself is defined as a condition where there is a linear relationship between several variables or all independent variables [8]. So it requires pattern recognition that can identify the target gas concentration in the sensor response data which has multicollinearity properties.

System Framework

The e-nose instrument is composed of hardware and software. The e-nose hardware is a set of electronic systems, where the aroma inputs are then transformed into output signals in the form of sensor responses. Meanwhile, the software is a system designed to further process sensor responses so that the system can identify the aroma of a given sample. In general, the overall picture is shown in the form of a block diagram in Figure 1. The aroma that comes from the sample is captured by the sensor array, then expressed in terms of the amount of resistance R (Ω). After that, the resistance value passes through the signal conditioning sequence for adjustment and is converted into an analog voltage V (mV). Adjustments are made so that the analog voltage can be converted into a digital voltage by the analog to digital converter (ADC). Furthermore, the sensor response which is already in the form of digital data through processing is displayed on the monitor screen and stored in the form of a text file [9].

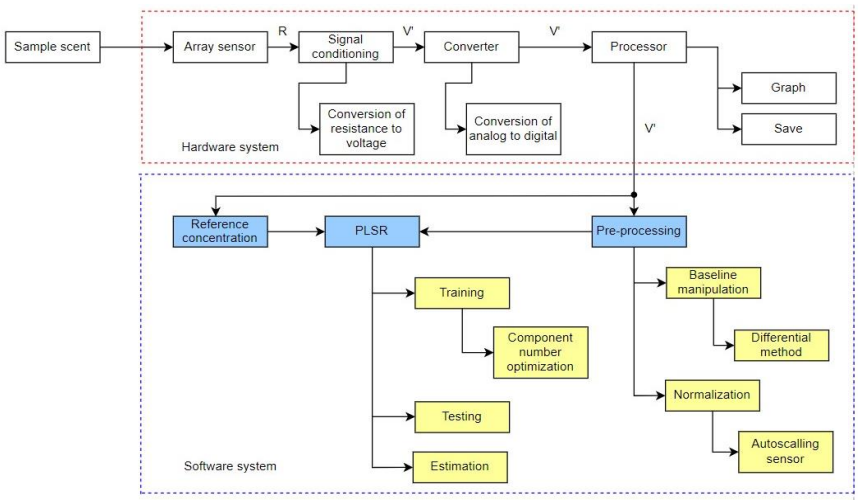


Figure 1 Block diagram of the instrument *e-nose*

The sensor responses that have been stored in the form of a text file are then analyzed further using a pattern recognition engine that is carried out through software. Digital raw data (raw data) is processed during preprocessing phase. At this time, unnecessary sensor response data are separated, so the stored data is aroma data for the collecting phase. Data preprocessing is carried out in two stages. The first stage is to refer to sensor data at the respective baseline. This process is carried out using a differential method in the baseline manipulation process. After the process, it is continued with the second preprocessing stage, namely normalization. Normalization is done to centralize the data in order to adjust the input for PLS regression (PLSR). Normalization is carried out using the autoscaling sensor method. The results of the normalization process are in the form of training data and test data (XN) which are ready to be used by PLSR. In addition to pre-processing, calculations are also carried out at this stage to find the reference concentration (Y) which acts as the response variable. The reference concentration was calculated using the response data from the TGS 832 sensor.

Regression analysis using PLSR was carried out by entering the data from a combination of normalized sensor response (X_N) and reference concentration (Y). After the training process, a training model is produced. The training model is a model that is built using the most optimal components. The resulting training model is then used to conduct testing. Tests are carried out to see the ability of the model to predict the input data that has never been studied. Then the best model is selected and used to predict real fluoride levels.

Pre- Processing

Analysis of the sensor response is preceded by preprocessing phase. In this study, two preprocessing stages were carried out, which are manipulating the initial value of the sensor and normalizing it. The first stage, baseline manipulation, is to process sensor responses to their respective baselines. Preprocessing utilizing baseline manipulation can mitigate sensor drift deviations while enhancing signal contrast and scalability. In this study, a differential method is used for the baseline manipulation stage. This method is executed by subtracting each sensor response value from its respective initial baseline values. As a result, sensor deviation (drift) within the sensor signal are effectively eliminated from the processed sensor response. Thus, the sensor response obtained is homogenized in term of its initial value. The second stage of preprocessing is normalization. The pre-processing technique used is an autoscaling sensor, known as standardization. This technique belongs to the global method used to ensure that the sensor magnitude is comparable and prevents the pattern recognition algorithm from being flooded by large value sensors. The normalization stage is given input in the form of a pre-processing sensor response and gives an output in the form of a normalized sensor response.

The Reference Concentration

The determination of fluoride reference concentration present in tea is not undertaken using GC-MS analysis, primarily due to the limitation associated with available GC-MS instrumentation in Indonesia. Another alternative to

get the reference concentration of fluoride in tea is done through another approach, namely using a formula to obtain the concentration (PPM) of the target gas attached to the sensor datasheet. The calculation of fluoride content in parts per million (ppm) based on the ratio of Rs/Ro is implemented for each response data of TGS832 sensor. Within this context, for every response dataset generated by TGS832 sensor, the calculation of the corresponding reference fluoride concentration is fine-tuned using the Ro parameter. The determination of Ro value is achieved by referencing the sensitivity characteristics graph of TGS832 sensor. This methodological approach is chosen based on the recognition that TGS832 sensor, as delineated in its datasheet, exhibits sensitivity towards gas compounds containing fluoride, specifically *chlorofluorocarbons*. The reference concentration, which will serve as the response variable (Y) for constructing the regression model, is determined through the utilization of Equation (1) as presented below:

$$\text{Concentration ref} = Rs / Ro \quad (1)$$

In Equation (1) the Rs parameter indicates the sensor resistance value at a certain gas concentration, while Ro is the resistance value when it is in clean air. The gas sensor provides output in the form of a digital voltage (mV), so it is necessary to convert from voltage (mV) to a resistance value. To calculate Ro, a value of Rs is needed when it is in clean air [11].

Independent Variables and Dependent Variables

The independent variable, or predictor variable, are denoted as X. The sensor response acts as a predictor. In this study, 12 sensor responses were used, so there were 12 predictor variables. The sensor response is measured in millivolts (mV). Meanwhile, the dependent variable, or response variable denoted as Y is the calculation of ppm fluoride based on sensor response (reference concentration). The reference concentration calculation refers to the sensor which has a sensitivity to fluoride-containing compounds. The independent variable consist of the sensor response data, while the dependent variables encompass the concentration derived from the approach of sensor response-based measurement. This indirectly shows the sensor response data has a dual role, namely as an independent and dependent variable. Where the dependent variable should be the concentration value of analytical chemistry measurement results such as GC-MS. The measurement of fluoride concentration cannot be done through GC-MS, due to the limitations of the instrument in Indonesia and high operational costs. Consequently, a middle ground is taken for the concentration value, namely through measuring the sensor response to the dependent variable. Even so, the sensor response is not necessarily a single component to get concentration, but there are other components to help calibrate the measurement, namely the sensor sensitivity characteristic graph issued by the factory and displayed through the sensor datasheet.

The Formation of PLSR Model

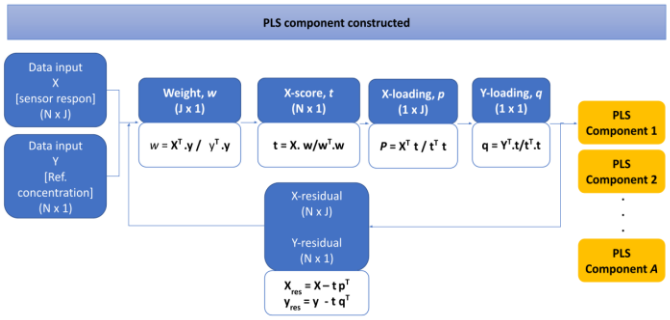


Figure 2 Block diagram of PLSR stages (modified from [12])

The stages of PLS regression (forming the PLSR model) to implementing the PLS model to make predictions can be seen in Figure 2. Figure 2 illustrates the procedure or stages of forming an A-PLS component using input X (sensor response) and input Y (reference concentration). The first step, the vector weight (w) is estimated by maximizing the covariance between x and y. Then, proceed with the sequential calculation and obtain the x-scores (t), x-loading (p), and y-loading (p). Then, in the end, the regression coefficient (b) is estimated using the w, p, and q that have been obtained, thus the first set of PLS components and the first loading have been successfully formed. Then the x-residual

(xres) and y-residual (yres) from the first PLS component become X data input and Y data input to build the second PLS component. This procedure is repeated up to A times, if an A-PLS component is required.

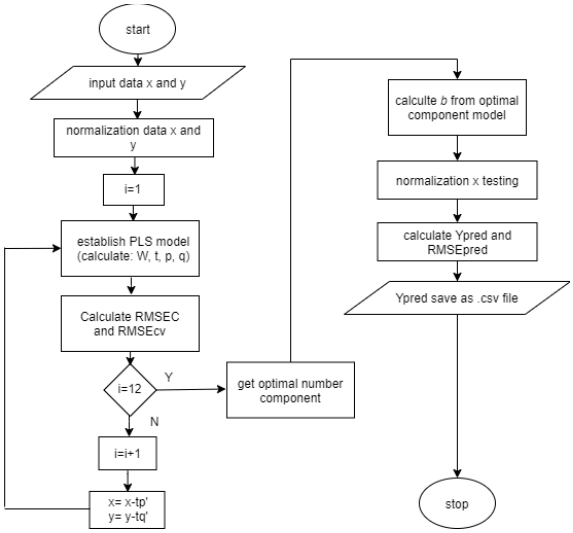


Figure 3 PLSR Flowchart

Figure 3 is a flowchart of prediction task using the PLSR method. The flow chart begins by initiating the input data that will be used to build the PLS model. Then the input data is normalized using the autoscaling sensor method. Furthermore, the PLS model is built through W, t, p, q. After the model is formed, the RMSEcv calculation is performed. The optimal number component is selected through the PLS model which produces a minimum RMSEcv value. Then the PLS regression coefficient (b) is calculated to calculate Ypred. Prior to conducting regression analysis with coefficient (b), a preliminary normalization is executed on the test data. Finally, the Ypred value that has been obtained is stored in a file with a .csv extension.

Prediction Using PLSR Model

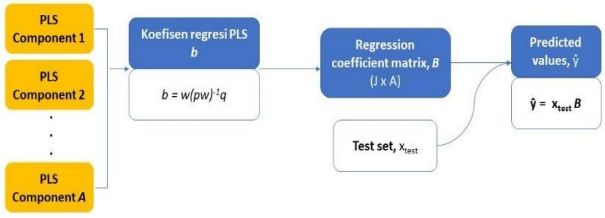


Figure 4 Block diagram of the prediction stage prediction

For predictive purpose, the testing set (xtest) is projected into a new dimension, specifically the PLS component through coefficient b, in order to generate predictive values (ŷ) or ypred. The process of conducting predictions using the PLS component is seen in block diagram (Figure 4). Each quality assessment of black tea is subjected to experimentation using PLSR across 20 combinations, yielding 20 regression models derived from the training process. Consequently, a comparative analysis is performed to select the most optimal model from the pool of 20 regression models that have been established.

Testing of the PLS model that has been formed is carried out using a validation set and a testing set. The estimation results using the validation set (x, y) are used to calculate the RMSEpred and R2pred, while the testing set (x) is used as the set to which the response variable (Y) will be predicted. The testing procedure was carried out for each data partitioning combination. The process of comparing the regression models involves evaluating the RMSEpred and R2pred values for each respective data partitioning combination.

Multicollinearity Test

The purpose of conducting multicollinearity testing is to ascertain whether there exists correlation among predictor variables within the regression model. One of the techniques employed to perform multicollinearity testing involves assessing the Pearson intercorrelation matrix (r). The Pearson correlation coefficient calculation was performed to assess the multicollinearity of the data [13]. The assessment of multicollinearity entails determining the Pearson correlation values among the predictor variables. As outlined by previous study [13], distinct correlation coefficient intervals exist, each with its own corresponding interpretation.

Electronic Nose

The e-nose instrument used in the previous study [9] consisted of 12 TGS sensors. The following are details of the sensors used in previous research which are detailed in Table 1. These sensors convert chemical information from various gases into electrical or measurable signals [14]. This conversion happens because the sensors respond to the concentration of specific particles in the gas. Using more sensors makes the electronic nose work better.

Table 1 Details of the sensors used [9]

Sensor	Description
MQ 7	Sensitive to CO gas, respond to other gases H2, LPG, CH4, alcohol, and air
TGS2600	Sensitive against pollutant gases, methane, CO, Iso Butane, Ethanol, Hydrogen, and air
TGS813	Sensitive to gas Methane, Propane, and butane, a flammable gas
TGS825	Sensitive to H2S low concentrations
TGS2602	Sensitive to H2S, Ammonia, Ethanol, Toluene
TGS826	Sensitive to Ammonia gas, response to Iso-butane, H2, Ethanol, and Air.
TGS2610	Sensitive to LPG, Propane, Iso-butane, Methane, Ethanol, H2, and air.
TGS 2611	Sensitive to Methane, Ethanol, Hydrogen, Iso-Butane, and Air.
TGS832	Sensitive to CFC (Chlorofluorocarbons) gas or Freon, ethanol, and air.
TGS2612	Sensitive to Methane gas, LP gas, Ethanol, Iso-Butane, Propane
TGS 2620	Sensitive to Alcohol, Hydrogen, IsoButane, CO, Methane, and air
TGS822	Sensitive to Alcohol, n-Hexane, Benzene, Methane, CO, Iso-Butane, Acetone.

The Taguchi Gas Sensor (TGS) is a type of gas sensor made of a metal oxide called SnO2 and produced by Figaro. TGS sensors are attractive due to their low cost, durability, and their ability to sense gases effectively. They are also known for their high sensitivity, quick response, low power consumption, and a simple design that doesn't take up much space [15]. The gas sensor material in TGS is SnO2, or tin dioxide. TGS is notably sensitive to the presence of oxygen, which can affect its resistance value. When there's more oxygen at the surface of the SnO2 layer, it raises the potential barrier, changing the sensor's resistance. When the SnO2 crystal is heated to a specific temperature, oxygen is attracted to its surface and becomes negatively charged. This is because the crystal's surface gives electrons to the outer-layer oxygen, making it negatively charged, while the outer surface of the crystal becomes positively charged. The difference in surface voltage then slows down the flow of electrons.

In this study, 12 sensor data inputs (predictor variables) and one target variable representing the calculated concentration of fluoride from the TGS 832 sensor were employed. All sensor data inputs were used due to the high multicollinearity among the data, while the target variable was derived solely from TGS 832 data because this sensor is sensitive to compounds containing fluoride as one of their constituent components.

RESULT AND DISCUSSION

Result of Multicollinearity Test

The primary objective of conducting a multicollinearity test is to examine whether there is correlation among predictor variables within a given regression model. The detection of multicollinearity within a dataset involves assessing the Pearson intercorrelation matrix of the predictor variables. By analyzing the Pearson correlation matrix, the presence of strong correlations (0.9 or higher) indicates potential multicollinearity. Illustrated in Table 2 is the Pearson correlation matrix representing the predictor variables, presented in a tabular format. Each variable, labeled as x1-x12, corresponds to 12 sensor response variables acting as predictors. Upon inspecting the Pearson correlation

matrix, a clear indication of multicollinearity among variables becomes evident, as the overall correlation values among predictor variables closely approach a value of 1. However, some variables also exhibit moderately sized correlations in comparison to other variables.

Tabel 2 Multicollinearity test results

Predictor	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
Xl	1.000000	0.940427	0.983293	0.881856	0.785425	0.869217	0.901821	0.896686	-0.056984	0.871875	0.894005	0.924602
X2	0.940427	1.000000	0.980909	0.974652	0.941613	0.930298	0.765414	0.727573	0.171496	0.701836	0.950423	0.985097
X3	0.983293	0.980909	1.000000	0.949774	0.972890	0.933179	0.849041	0.836277	0.033529	0.804254	0.951244	0.976352
X4	0.881856	0.974652	0.949774	1.000000	0.945372	0.973676	0.711235	0.661022	0.174703	0.642842	0.986246	0.993019
XS	0.945097	0.941613	0.972890	0.945372	1.000000	0.974214	0.847833	0.829118	-0.071181	0.811039	0.966841	0.965860
X6	0.869217	0.930298	0.933179	0.973676	0.974214	1.000000	0.746108	0.706371	0.032541	0.693896	0.985487	0.970400
X7	0.901821	0.765414	0.849041	0.711235	0.847833	0.746108	1.000000	0.903053	-0.413704	0.991671	0.766973	0.761626
X8	0.896686	0.727573	0.836277	0.661022	0.829118	0.706371	0.903053	1.000000	-0.283357	0.914245	0.717676	0.732107
X9	-0.056984	0.171496	0.033529	0.174703	-0.071181	0.032541	-0.413704	-0.283357	1.000000	-0.487539	0.053001	0.133571
X10	0.871875	0.701836	0.804254	0.642842	0.811039	0.693896	0.991671	0.914245	-0.487539	1.000000	0.710892	0.700440
X11	0.894005	0.950423	0.951244	0.986246	0.966841	0.985487	0.766973	0.717676	0.053001	0.710892	1.000000	0.986171
X12	0.924602	0.985097	0.976352	0.993019	0.965860	0.970400	0.761626	0.732107	0.133571	0.70044	0.986171	1.000000

Results of PPM Measurement Through TGS-832 Sensor

Each sensor is required to provide its sensitivity characteristic graph, which is usually furnished by the manufacturer and presented in the respective sensor datasheets. In the sensitivity characteristic graph, the x-axis represents the concentration in mg/L, while the y-axis represents the ratio Rs/Ro. In the context of this study, the ratio value plays a pivotal and principal role in determining the appropriate concentration values based on the available datasheets. Nonetheless, there has been no further research conducted to measure the accuracy of the Rs/Ro and mg/L values [10]. The gas sensor TGS-832 operates with a voltage source of 5 volts DC, and the voltage value emitted by the TGS-832 sensor is the reference utilized during the calibration process for converting it into parts per million (PPM). The calibration procedure for the gas sensor TGS-832 employs information documented in the datasheet. The datasheet provides information regarding the Load Resistance (RL) value designated for the TGS-832 sensor, which is indicated as 10k ohms.

The datasheet for the TGS832 sensor includes a graph that illustrates the relationship between ppm (parts per million) and Rs/Ro. Examining the graph reveals that the resistance of fresh air remains constant. Consequently, based on the graph, an estimated value of Rs/Ro can be obtained, specifically Rs/Ro = 2,45. The datasheet provides the formula for calculating the value of Rs for fresh air, as presented in Equation (2) as below:

$$Rs_{air} = \frac{V_c \cdot R_L}{V_{RL}} - R_L \quad (2)$$

Using formula on Equation (2), the values of Vc (source voltage, +5V) and RL (Load Resistance) from the datasheet are employed, with RL being 10k ohms. Subsequently, VRL represents the voltage output value during the sensor response in the flushing state. Utilizing Equation (2) yields an Rsair value of 46078,70 ohms. Consequently, the obtained Ro value is 1881,55 ohms. This establishes that sensor readings to determine ppm can be achieved through the Rs/Ro relationship. The readings from the sensor for obtaining fluoride ppm are depicted in Figure 5.

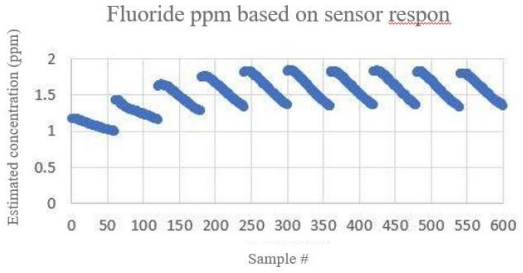


Figure 5 The outcomes the relative concentration of fluoride

In Figure 5 a scatter plot is shown from the calculation of ppm fluoride based on the sensor response for kw2 black tea samples. Based on the plot, there are 10 scatter lines, which represent 10 collecting phases. Each scatters line goes downhill as it enters the purging phase, where the sample aroma is purged from the sensor chamber. If you look at the lines formed from the scatter plot, the lines look like a logarithmic graph, but they run downhill because they are blocked by the purging phase. In the TGS-832 datasheet, the sensitivity graph is presented in a logarithmic graph. Based on Figure 5, the ppm reading of fluoride via the Rs / Ro ratio sensor increases with the length of collecting time. The results of this reading will then be used as the Y response variable to build a predictive model for fluoride concentration using the PLSR method.

Analysis of The Prediction Results

The predictions generated by the Partial Least Squares (PLS) model using the test set data as input are saved in the form of .csv files. The assessment indicators for the predictive performance of each PLS model are quantified through the values of R2, RMSE, R2cv, RMSEcv, R2pred, and RMSEpred. These metrics are presented in Table 3.

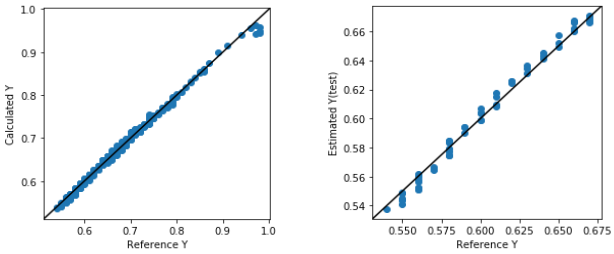
Black tea as a drink that is high in fluoride (fluoride) stores fluoride not in the single form of the elemental fluoride, but in various forms. The fluoride element in tea is stored in the form of water-soluble form, residual-F, organic-bound F, iron-manganese-bound, exchangeable F, and PAHs. So that the amount of fluoride contained in tea carried out in this study is the amount of fluoride that can be detected by the instrument e-nose.

The prediction process employs Partial Least Squares Regression (PLSR) to estimate the fluoride quantity in samples in units of parts per million (ppm). The predictor variable set, denoted as X in PLSR, comprises responses from 12 sensors. In contrast, the response variable set, denoted as Y in PLSR, originates from concentration measurements via the TGS832 sensor response approach. Notably, variable selection was omitted in this study due to the multicollinearity present in the sensor response data. Nevertheless, the PLS regression results indicate an absence of redundancy that could compromise prediction accuracy, owing to the abundant information provided by the predictors. This is evident from the satisfactory coefficient of determination of prediction (R2pred) values, as depicted in Table 3. Furthermore, this underscores the robustness of PLSR as a regression technique in the presence of multicollinearity within the data.

Tabel 3 The results of assessment indicators on the predictive ability of model

	KW 1		**KW 2**		**KW 3**	
	70% train	**80% train**	**70% train**	**80% train**	**70% train**	**80% train**
Component	8	8	5	5	7	8
PRESS	0.199926	0.207172	1.858669	1.692461	0.431934	0.427480
R^2	0.994960	0.994932	0.989050	0.989689	0.986699	0.987830
RMSEC	0.005642	0.005509	0.024101	0.023338	0.012210	0.011452
R^2cv	0.984930	0.985067	0.844944	0.887026	0.960374	0.964675
RMSEcv	0.009755	0.009456	0.090694	0.077250	0.021076	0.019512
R^2pred	0.985840	0.986620	0.985573	0.994791	0.993024	0.994054
RMSEpred	0.004907	0.004642	0.019185	0.010772	0.006446	0.005324

The chosen e-nose data, which serves as the regression model derived from the training phase, must adhere to criteria encompassing R2, RMSEC (Root Mean Square Error of Calibration), RMSEcv (Root Mean Square Error Cross Validation), and RMSEpred (Root Mean Square Error of Prediction) values. RMSEC is utilized to assess the presence of errors within the training model. Subsequently, the accuracy of the training model is tested through cross-validation. Cross-validation techniques are applied to predict or estimate the accuracy of the training model for implementation. RMSEpred serves as one of the metrics employed to validate the model, with lower values signifying a more valid model [16].



(a) (b)

Figure 6 (a) Plot Y with Ŷ for the training model (b) and the predicted results kw1

Figure 6 illustrates the linearity of the relationship between predicted fluoride values and reference fluoride values in kw1 for the 80% training data. Based on Table 3, the root mean square error of cross-validation (RMSEcv) is calculated as 0.009456, and the root mean square error of prediction (RMSEpred) is calculated as 0.004642. When comparing the PLS models for kw1 using 70% and 80% training data, the model trained with 80% data demonstrates superior predictive capabilities, exhibiting an RMSEpred value of 0.004642 and an R2pred value of 98.6%. For the black tea sample kw1, the coefficient of determination for prediction (R2pred) of 0.986620 indicates that 98.6% of the variance in the reference concentration (Y) can be elucidated by the variability in the gas sensor responses (X). This robustly signifies a strong positive linear relationship between the two variables, as supported by the regression model, which indicates that the variability in the reference concentration can be predicted using gas sensor responses with a magnitude of 98.6%.

The activity of making predictions with many predictors is a complex activity compared to predicting a value using one predictor. Because in this case, predictions with multiple predictors need to consider the error limit. The more precise the prediction, the smaller the error. The coefficient of determination (R2) in this context represents the size of the error, with the maximal measure of how much variation in the predictor can be explained by the resulting model, to predict the response. In Table 3, the predictive determination coefficient value for the three tea samples gives satisfactory results, 12 sensor response data are also used in which there are redundancy and irrelevant variables. Based on the models with the best predictive ability for each tea quality, the average predicted fluoride concentrations in each sample can be seen in Table 4. Based on this table, it can be seen that the three samples are still within the allowable threshold. In addition, research conducted by [7] measured the concentration of fluoride in Iranian black tea using GC-MS of 1.09 ppm. Based on the research data, the prediction of fluoride concentration in kw3 black tea is close to the results of previous research [7], with a difference of 0.3 ppm.

Tabel 4 Average forecast fluoride concentration in black tea

	KW1	**KW2**	**KW3**
Average prediction (ppm)	0.676936	1.357968	0.803479

CONCLUSION

We have successfully designed a software system that can be used to predict the concentration of fluoride content through an approach reference Rs / Ro on the aroma of three types of black tea with different qualities using the PLSR method. The predictive performance value obtained based on the reference concentration (Rs / Ro) for each black tea sample quality was satisfactory.

It is possible to enter the fluoride concentration measured using GC-MS as the reference concentration in the training data to optimize the predicted results. The PLSR method can be implemented to convert aroma to a target concentration in other cases.

REFERENCES

1. Y. Zhang, X. Huang, S. Liu, W. Qian, X. Yang, and Y. Chen, "The forms of fluorine in tea and its bioaccessibility," vol. 52, no. July, pp. 289–298, 2019. Available at:. www.fluorideresearch.org/epub/files/028.pdf

2. U. Zulfaida, A. Arumsari, and R. Nugraha, "Analysis of Fluoride Levels in Steeping Tea Samples (Camellia Sinensis) using the Visible Light Spectrophotometric Method," vol. 2, no. 2460–642, pp. 15–22, 2016. Available at: http://karyailmiah.unisba.ac.id/index.php/farmasi/article/view/3040
3. J. Palacín *et al.*, "Application of an array of metal-oxide-semiconductor gas sensors in an assistant personal robot for early gas leak detection," *Sensors (Switzerland)*, vol. 19, no. 9, pp. 1–16, 2019. Available at https://doi.org/10.3390 / s19091957
4. I. Inca, TW Widodo, and D. Lelono, "Classification of Tambi-Pagilaran Green and Black Tea using the Principal Component Analysis Method (PCA) Using E-Nose, ". Indonesian Journal of Electronics and Instrumentation Systems (IJEIS), vol. 8, no. 61, 2018, pp. 2460-7681. Available at https://doi.org/10.22146/ijeis. 28718.
5. L. Suryani, "Classification of black tea based on e-nose using independent component analysis and SVM," Yogyakarta, 2019.
6. MJ Azzuhdi, "Classification of black tea based on planting location using e-nose with a supervised method learning, "Yogyakarta, 2020.
7. M. Khiadani (Hajian)," Determination of polycyclic aromatic hydrocarbons concentration in eight brands of black tea which are used more in Iran, " *Int. J. Environ. Health Eng.*, vol. 2, no. 1, p. 40, 2013. Available at https://doi.org/10.4103 / 2277-9183.122427
8. WA Puspita, "Comparison of Partial Least Square (Pls) Methods And Principal Component Regression (Pcr) To Overcome Multicollinearity In Multiple Regression Models," State Islamic University of Sunan Kalijaga, 2018. Available at http://digilib.uin-suka.ac.id/id/eprint/32140
9. D. Lelono, "Development of Electronic Nose System Instrumentation for Local Black Tea Test," Universitas Gadjah Mada, 2017. Available at http://etd.repository.ugm.ac.id/pen
10. Research/detail/110036SI Sabilla, R. Sarno, and J. Siswantoro, "Estimating Gas Concentration using Artificial Neural Network for Electronic Nose, " *Procedia Computer Science.*, vol. 124, pp. 181–188, 2017. Available at https://doi.org/10.1016/j.procs.2017.12.145
11. KR Zigner and B. College, "Methane Sensing in the Field", A Senior Project submitted, 2019. Available at https://digitalcommons.bard.edu/senproj_s2019/229
12. LC Lee, CY Liong, and AA Jemain, "Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: A review of contemporary practice strategies and knowledge gaps, " *Analyst*, vol. 143, no. 15, pp. 3526–3539, 2018. Available at https://doi.org/10.1039 / C8AN00599K
13. S. Yadav, "Correlation analysis in Biological Studies," *J. Pract. Cardiovasc. Sci.*, Vol. 3, no. May, pp. 36–8, 2018. Available at https://doi.org/10.4103 / jpcs.jpcs_31_18
14. Anggriawan, R. Y.,"Studi Penentuan Kadar Nikotin pada Rokok Elektronik menggunakan E-Nose dengan kemometrik sebagai analisisnya", Departemen Fisika, Universitas Gadjah Mada, (2019)
15. Faiz, Suyanto. ST, M., "Uji Kinerja Taguchi Gas Sensor (TGS) untuk Monitoring Gas Methane pada Portable Biodigester", Jurusan Teknik Fisika FTI ITS Surabaya, (2012)
16. M. Ohyver, "Penerapan Partial Least Square Pada Data Gingerol", Matematika dan Statistik Fakultas Sains dan Teknologi, Universitas Bina Nusantara, Juni (2010)