Stability Assessment of Feature Selection Algorithms on Homogeneous Datasets: A Study for Sensor Array Optimization Problem

DEDY RAHMAN WIJAYA [ID]1 AND FARAH AFIANTI [ID]2

[1]School of Applied Science, Telkom University, Bandung 40257, Indonesia
[2]School of Computing, Telkom University, Bandung 40257, Indonesia

Corresponding author: Dedy Rahman Wijaya (dedyrw@tass.telkomuniversity.ac.id)

ABSTRACT A feature selection algorithm (FSA) is used to eliminate redundant and irrelevant features. Obviously, it can reduce dimensionality as well as the complexity of the original problem. Furthermore, the stability of FSA output becomes a major issue in real-world applications. Stability refers to the consistency of its feature preference related to the perturbation of data samples. In sensor array optimization, an FSA is used to find the best sensor combination in a sensor array. Typically, the main objectives of sensor array optimization are reducing data dimensions, electrical consumption, production cost, computational and traffic overhead, etc. Furthermore, the stable outputs of FSA in several observations are necessary to make a firm conclusion of selected sensors. The contribution of this research is to investigate the stability of FSAs in twelve homogeneous datasets in relation to the sensor array optimization problem. In this study, the stability of seventeen filter-based FSAs is compared across twelve homogeneous datasets. These datasets are generated from the electronic nose (e-nose) used to monitor twelve types of beef cuts. In this case, gas sensor array must have good generalization to differentiate all beef types. The experimental results show that a single FSA cannot guarantee stable sensors recommendation in sensor array optimization. Thus, it becomes a caution to researchers and practitioners to use a proper approach when performing sensor array optimization.

INDEX TERMS Electronic nose, feature selection, sensor array optimization, stability assessment.

I. INTRODUCTION

Gas sensor array is the main component of the e-nose that detects and collects volatile information from a particular object [1]. It is assembled from several gas sensors with different selectivity. The sensor combination in the sensor array depends on an application of e-nose to detect a particular sample. The different sample has different volatiles that acts as biomarkers. For example, e-noses have been used to various applications such as blood glucose level detection [2], halal authentication [3], [4], meat quality detection [5]–[10], classifying vegetable oils and animal fats [11], tea classification [12], [13], monitoring tempeh fermentation [14], etc. In e-nose application, not optimal sensor array susceptible to overlapping selectivity which means

more than one sensor has similar selectivity with others. Hence, sensor array optimization is performed to reduce this problem. Moreover, several advantages can be obtained by performing sensor array optimization such as performance improvement, reducing data and communication overhead, saving electrical consumption, reducing production costs, etc [15]. In a sensor array optimization problem, sensor combination is not only optimal but also has good generalization. Typically, FSA is employed to determine the best sensor combination in a sensor array. Furthermore, the generalization of sensor array only can be guaranteed by using a stable FSA based on several datasets.

FSA plays an important role to reduce high-dimensional data for classification and regression tasks. Reducing data dimension can give several advantages such as computational efficiency, lower data storage, simpler data model, lower training time, etc. Several approaches have been proposed

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Tan [ID].

D. R. Wijaya, F. Afianti: Stability Assessment of FSAs on Homogeneous Datasets: Study for Sensor Array Optimization Problem

IEEE *Access*

for feature selection including filter, wrapper, and embedded. Filter-based FSAs typically rank or select features by considering relation or correlation between features and class label. The advantages of this algorithm are fast, independent of the classifier, lower computational complexity, robust against overfitting, and better generalization. Because filter-based FSAs are independent of learning algorithms, they can fail to select the best feature combination in classification and regression tasks [16]. In contrast, wrapper intensively interacts with a classifier that makes it can find the best feature combination for a particular learning algorithm. However, it requires computationally intensive and susceptible to overfitting. Embedded FSA also relies on the classifier to determine the best feature subset even though it has better computational complexity than wrapper FSA [17].

Stability and generalization have become major issues in FSA studies. The stability of FSA indicates the robustness of its feature preference, related to the perturbation of data samples [18]. Instability of selected features leads to the difficulty to make a conclusion and lack of feature set generalization. There are several causes of FSA instability including small sample size, sample order dependency, and data partitioning [19]. In this study, the stability of FSAs is assessed based on homogeneous datasets that is generated by e-nose in beef quality monitoring. The objective of this experiment is to solve sensor array optimization problem which emphasizes the robustness of FSAs, so sensor combination is not only optimal but also has good generalization to detect various types of beef samples. In this study, we have several motivations as follows:

1. In the existing studies, sensor array optimization in e-nose employs a single FSA over a limited dataset which is a common approach to deal with this problem. However, to the best of our knowledge, there is no study that addresses the stability of FSAs in sensor array optimization across several datasets.

2. Several studies investigate FSA stability in several types of data including high-dimensional spaces [20], high-dimension and correlated data [21], high-dimension with small sample data [19], [22]. However, the majority of these studies use artificial data. On the other hand, in this study, we use homogeneous datasets that are generated by e-nose in the real case of beef quality monitoring. Homogeneous dataset refers to the datasets obtained from different observations in the similar environment. Furthermore, the stability of FSA becomes a major issue in sensor array optimization. Thus, it can be considered as new and important in this area.

Hence, the main contribution of this research is to investigate the stability of FSA in twelve homogeneous datasets in relation to the sensor array optimization problem.

The rest of this paper is organized as follows: Section II discusses related works including why sensor array optimization is important, the existing studies of sensor array optimization, and FSA stability issues. Section III is the problem formulation that explains the experimental setup, feature selection algorithms, weighting, and stability metric. Section IV discusses the result of the stability assessment. Finally, Section V draws a conclusion.

II. RELATED WORKS

The gas sensor array is a combination of several gas sensors that consist of different gas selectivity. Each sensor works individually and simultaneously to convert chemical information from various gases into a measurable signals. This condition causes overlapping selectivity which one or more gases can be detected by more than one gas sensors. The main problem arises in selecting the best and most effective gas sensor combination. Gas sensors need high power consumption especially for wireless communication than for sensing [23]. The gas sensor technology that consumes the highest power is the optical gas sensor platform therefore it is not suitable to be implemented in the Wireless Sensor Network (WSN). On the contrary, catalytic and semiconductor-based sensors provide a balanced trade-off among three-parameter such as power consumption, safeness, and performance [23]. Besides gas sensor technologies, the other ways to optimize power consumption are power management, sensing circuits, and measurement procedure.

Sensor array optimization is one of the effective mechanisms to deal with the power consumption issues in the gas sensor nodes. In 2007, sensor array optimization to detect the quality of wheat using ANOVA followed by multiple comparisons (Tukey) was proposed [24]. This method can reduce the number of gas sensors from 10 to 5. Furthermore, the percentage of correct classification increases by about 8% than without optimization. Two years later, the lowest sub-arrays to detect 11 variants of gasses were proposed [25]. Several gas sensors are clustered to fill each level of smaller sub-arrays. The number of gas sensors reduces about 50% from 6 to 3 gas sensors. In 2010, the combination among cluster analysis (CA) for preliminary and genetic algorithms (GA) for decision are proposed [26]. The datasets consist of three different resources. Two among them obtained from the real experiment by using 5 sensors equipped with 7 vapors and 10 sensors equipped with 3 vapors. The optimum number of sensors in both datasets is reduced into 2 sensors. Moreover, the simulation version of datasets using 10 sensors equipped with 3 vapors is reduced into 4 sensors. The general resolution factor (GRF) increases ranging from 0.3 up to 8.5%. One year later, decision methods are improved using genetic algorithms for multi-objective optimization [27]. Furthermore, the number of datasets increases from 3 to 4. The simulation dataset that uses 10 sensors and 5 vapors are added. The number of vapors from the last experiment is redefined. The optimum number of the sensor from datasets 1, 2, 3 and 4 are 3, 3, 2 and 2 sensors, respectively. Moreover, GRF from the selected sensor is better than the value of all sensors or the value of the partial random sensor. In the same year, the neural network sensitivity analysis for volatile

IEEE *Access*

D. R. Wijaya, F. Afianti: Stability Assessment of FSAs on Homogeneous Datasets: Study for Sensor Array Optimization Problem

organic compound (VOC) mixture was proposed [28]. The number of optimum sensors reduces from 6 to 4. In addition, the experiment result shows that the number of the sensor to detect the gas mixture is the lowest. In the other case, the rough set-based approach to classify the quality of black tea was proposed [29]. The number of optimum sensors reduces from 8 to 4. Furthermore, the separability index (measurement to show the fraction of data point that has the same labels as the closest neighbor) and the level of accuracy increase about 3% and 11%. In 2012, the same case was classified using the t-score, fisher's criterion, and minimum redundancy maximum relevance (MRMR) [30]. The number of optimum sensors reduces from 8 to 3. Moreover, the level of accuracy increases ranging from 6 to 10%. Two years later, the integration of genetic algorithm (GA) and quantum-behaved particle swarm optimization (QPSO) to detect wound infection was proposed [31]. The optimization uses a weighting between 0-1 but the actual number of sensors does not change that is 15 sensors. The advantage of this method is the level of accuracy increases by about 7.5%. In the same year, 2014, kernel principal component analysis (KPCA) based linear discriminant analysis (LDA) to detect indoor air contaminants was proposed [32]. This method can reduce a sensor from 4 sensors. High performance followed by low-cost implementation of air contaminants can be detected using the combination of 3 gas sensors. In 2016, binary quantum-behaved particle swarm optimization to detect wound infection was proposed [33]. This method can reduce the number of sensors from 20 to 6 sensors. Furthermore, the level of accuracy reaches 97.6%. In the same year, the wavelet transform and filter-based feature selection approach to classify beef quality was proposed [15]. However, this study only used single FSA and single dataset. One year later, an optimized feature matrix using mean, variation coefficient, cluster, and correlation analysis to detect Chinese pecan quality was proposed [34]. The number of the sensor was reduced by 5 sensors from 13 sensors. Moreover, the data dimension was reduced by 19 from 30. The result was shown in the principal component analysis (PCA) score plot and regression model. In the same year, 2017, the integration of traditional ANOVA including loading analysis methods, Wilks statistic method and sensors sensitive to aroma compounds for detecting a variety of apple juices was proposed [35]. The results were tested using PCA, K-means clustering, and SVM. The number of the original sensors was 10. The optimization algorithm using ANOVA including loading analysis methods, Wilks statistic method, and sensors sensitive to aroma compounds can reduce the number of sensors into 7, 4 and 7, respectively. SVM combined with the Wilks statistic method reaches 100% of testing accuracy. Furthermore, the lowest number of the sensor increases efficiency and flexibility for a large number of apple juice samples. One year later, 2018, random forest completed with new measurement namely Gini importance was proposed [36]. The number of sensors was reduced from 6 to 2 sensors based on their best accuracy that was predicted

before using Gini importance. In 2019, the response surface method (RSM) was proposed to optimize the number of the sensor in detecting the freshness of strawberry [37]. The number of sensors was reduced from 8 to 5. The classification methods to analyze the result were PCA, LDA, and SVM. The accuracy of LDA classification was 86.4% while the data variance explanation for LDA was 84%. The validation accuracy of two types of SVM was 50.6% for C-SVM and 55.6% for Nu-SVM.

Typically, in sensor array optimization, FSA is employed to select the best feature set related to gas sensor combination. However, all of the previous studies were not discuss about robustness and stability of the FSA in the sensor array optimization. Practically, FSA can generate unstable output on the homogeneous dataset. This problem leads to lack of generalization for sensor array and difficulty to determine which the best gas sensor combination. Moreover, several studies concern on FSA stability issues. The stability of five FSAs including are assessed on a set of proteomics datasets [20]. In particular, the stability of wrapper FSA is measured over four datasets and k-Nearest Neighbor is used as a classifier [38]. This study mentions that different training data partitioning leads to different selected feature set. Data heterogeneity also affects the ranking and the stability of FSA in high-dimensional correlated data [21]. Another study defines properties of stability measures including fully defined, bound, maximum, and correction of chance to assess several stability measures [39]. An assessment of FSA stability is also performed on high-dimension and small sample data [22]. This study concludes that small sample data is highly affected on FSA stability. FSA stability is also an issue in software quality prediction datasets [40]. The result shows that filter-based FSA (ReliefF) is more stable than wrapper-based FSA.

III. PROBLEM FORMULATION
In this section, the experimental setup to generate datasets is discussed. Moreover, the problem is mathematically formulated and the metrics of FSA stability is also explained

A. EXPERIMENTAL SETUP
In this experiment, twelve kinds of beef cuts were observed including round (shank), top sirloin, tenderloin, flap meat (flank), striploin (shortloin), brisket, clod/chuck, skirt meat (plate), inside/outside, rib eye, shin, and fat. Each of them is weighted 125 grams. Table 1 shows all of gas sensors that used in this experiment. Furthermore, prototype of e-nose sensor box equipped by wireless communication module is shown in Fig. 1. E-nose signals are sent from the sensor box to the computer server every minute over the wireless network. For each experiment cycle, data is recorded continuously for 2220 minutes from fresh beef to spoil. After one experiment cycle is complete, the temperature control box and sensor box are flushed using a high speed fan. After that, they were rested for 3-6 hours to neutralize the remaining odor from previous experiments. This procedure is repeated for all cuts of beef.

D. R. Wijaya, F. Afianti: Stability Assessment of FSAs on Homogeneous Datasets: Study for Sensor Array Optimization Problem

IEEE Access

TABLE 1. List of used gas sensors.

Gas sensor	Selectivity
MQ135	Ammonia, carbon dioxide, alcohol, benzene, smoke, NOx
MQ136	Hydrogen sulfide
MQ137	Ammonia
MQ138	Aldehydes, alcohols ketones
MQ2	Methane, alcohol, liquefied petroleum gas (LPG), hydrogen, smoke, propane, i-butane
MQ3	Alcohol, benzine, methane, hexane, LPG, carbon monoxide
MQ4	Methane
MQ5	Hydrogen, LPG, methane, carbon monoxide, alcohol
MQ6	Propane, LPG, iso-butane
MQ8	Hydrogen
MQ9	Methane, carbon monoxide, and propane
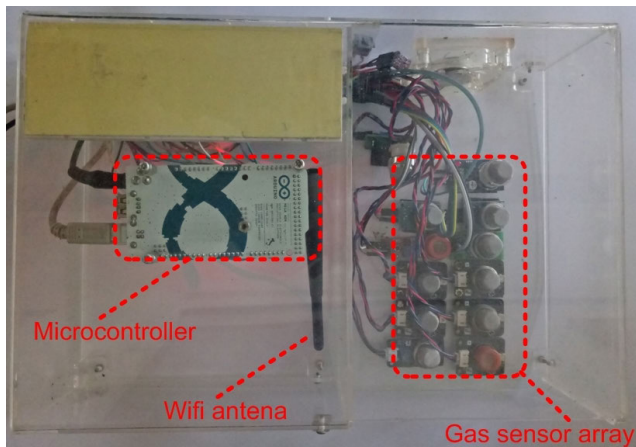
FIGURE 1. E-nose hardware prototype.

In one experiment cycle, we got 2220 measurement points from a beef cut. Thus, we have a total of 26640 measurement points from twelve datasets corresponding to twelve pieces of beef. For labeling, the total number of bacteria is used as main standard of beef quality. Spectrophotometer with 1000x dilution is used to quantify optical density. Afterwards, hemocytometer was utilized to determine the microbial population in a beef sample. The experiment refers to the combination of classical and two-hour method [41]. The beef quality is divided into four sensory classes according to total viable count (TVC). It complies with the standard issued by the Agricultural and Resource Management Council of Australia and New Zealand as shown in Table 2 [42]. According to this trait, they can consider as homogeneous datasets because they have almost similar pattern except the noise contamination produced by fluctuating humidity levels. Beside the common cause of small sample size, the stability is also highly dependent on the types of feature selection algorithm in use [19]. In this experiment, we used datasets generated by e-nose in beef quality monitoring. The main characteristic of these datasets is homogeneous, noisy, and relatively low dimension. The homogeneous dataset refers to the datasets obtained from different observations in the similar environment. Moreover, noisy data is caused by fluctuating relative

TABLE 2. The beef quality standard.

Class	TVC(log10 cfu/g)
Excellent	<3
Good	3-4
Acceptable	4-5
Spoiled	>5

*cfu/g : colony forming unit of bacteria in a gram of meat

humidity in the sample chamber. The issue of noise has been tackled by our proposed noise filtering framework, the interested reader can refer to [8], [9]. The number of used sensors is affected to the data dimension. In this experiment, eleven gas sensors produce eleven features which are relatively low. However, in sensor array optimization problem, the number of sensor in sensor array becomes a sensitive issue because the more sensors lead to increasing electrical consumption, production cost, and data traffic/storage. Generally speaking, the datasets have a total of 26640 measurement points from twelve datasets corresponding to twelve pieces of beef. We argue that the size of datasets is sufficient to deal with a small dataset issue. Moreover, it is not a good idea to use different sensor combination for different beef types that lead to a huge number of sensors utilization to build a sensor array. Thus, stability assessment of FSA is necessary to deal with sensor array optimization problem.

B. FEATURE SELECTION ALGORITHMS AND WEIGHTING
In this study, we investigate stability of 17 filter-based FSAs for sensor array optimization in e-nose. The brief explanation about FSAs including selection criterion is shown in Table 3. In this experiment, we used FEAST toolbox for information-theoretic FSAs [47] and scikit-feature for similarity-based and statistical-based FSAs [49]. Every FSA has different way to select the best features. There are two most important aspects of feature selection such as maximizing relevancy and minimizing redundancy. Relevancy means that the selected features must have ability to predict class labels. On the other side, redundancy indicates that the selected features should not have strong correlation with each other. It implies that several features with strong correlation can be represented by only one feature. Hence, in this subsection, the mechanisms of FSAs are briefly explained for better understanding in relation to stability assessment in the next section.

Consider dataset $DS_i = \{(x_j, c_j), j = 1, \ldots, n\}, i = 1, \ldots, m$, it consists of m and n series and instances correspond to number of beef samples and measurement points, respectively. Instance x_j denotes a k-dimensional vector $x_j = (x_{j1}, x_{j2}, \ldots, x_{jk})$ labeled as c_j where each component x_j represents the value of feature vector. Furthermore, consider an FSA whose output is a vector y that denotes a following selected feature subset

$$y = (y_1, y_2, y_3, \ldots, y_k), \quad y_i \in \{0, 1\}, \quad (1)$$

where a selected feature set is determined by a weighted appearance of a particular feature x_{jl} in dataset $DS(w_{DS_i, x_{jl}})$

IEEE Access

D. R. Wijaya, F. Afianti: Stability Assessment of FSAs on Homogeneous Datasets: Study for Sensor Array Optimization Problem

TABLE 3. Filter-based FSAs used in the experiment.

No	FSA	Description and selection criterion(score)
1	Mutual information maximization/information gain (MIM)	MIM is the simplest information-theoretic feature selection that has selection criterion for feature X_k related to label Y as follows: $\jmath_{MIM}(X_k) = I(X_k; Y)$. It only contains one term to determine the relevancy between feature and class label without considering features redundancy [43]
2	Mutual information feature selection (MIFS)	To deal with redundancy between feature X_k and X_j, MIFS adds the second term $\beta \sum_{x_j \epsilon S} I(X_k; X_j)$. Hence, the scoring can be expressed by $\jmath_{MIFS}(X_k) = I(X_k; Y) - \beta \sum_{x_j \epsilon S} I(X_k; X_j)$ [44].
3	Minimum redundancy maximum relevance (MRMR)	MRMR modifies the β value to be dynamically changed according to the number of selected features in feature subset S . The more number of selected features imply the effect of feature redundancy is gradually decreased. Thus, the selection criterion is formulated by $\jmath_{MRMR}(X_k) = I(X_k; Y) - \frac{1}{|S|} \sum_{x_j \epsilon S} I(X_k; X_j)$ [45].
4	Conditional infomax feature extraction (CIFE)	CIFE has the third term to consider conditional redundancy between unselected and already selected features given class labels ($\sum_{x_j \epsilon S} I(X_j; X_k|Y)$). This third term should be maximized, so the score is calculated by $\jmath_{CIFE}(X_k) = I(X_k; Y) - \sum_{x_j \epsilon S} I(X_k; X_j) + \sum_{x_j \epsilon S} I(X_j; X_k|Y)$ [46].
5	Conditional mutual information (CMI)	CMI considers the value of mutual information that is shared by feature X_k and labels Y given already selected feature subset X_S that can be expressed as follows: $\jmath_{CMI}(X_k) = I(X_k; Y|X_S) = I(X_k; Y) - \beta \sum_{x_j \epsilon S} I(X_k; X_j) + \lambda \sum_{x_j \epsilon S} I(X_j; X_k|Y)$, where β and λ are two non-negative parameters between 0 and 1 [47].
6	Joint mutual information (JMI)	JMI has principle to include new features that complement already selected features. It can be expressed by the condition likelihood maximization framework [47] as follows $\jmath_{JMI}(X_k) = \sum_{x_j \epsilon S} I(X_j; X_k|Y) = I(X_k; Y) - \frac{1}{|S|} \sum_{x_j \epsilon S} I(X_k; X_j) + \frac{1}{|S|} \sum_{x_j \epsilon S} I(X_j; X_k|Y)$ [48].
7	Conditional mutual information maximization (CMIM)	CMIM selects features by minimizing feature redundancy with maximizing this deduction factor ($\max_{x_j \epsilon S}[I(X_j; X_k) - I(X_j; X_k|Y)]$). Hence, the selection criterion can be expressed by $\jmath_{CMIM}(X_k) = I(X_k; Y) - (\max_{x_j \epsilon S}[I(X_j; X_k) - I(X_j; X_k|Y)])$ [49].
8	Double input symmetrical relevance (DISR)	DISR utilized normalized mutual information as criteria ($\jmath_{DISR}(X_k) = \sum_{x_j \epsilon S} \frac{I(X_j X_k; Y)}{H(X_j X_k; Y)}$) [50]. In another words, DISR is normalized version of JMI.
9	Interaction capping (ICAP)	The selection criterion of ICAP is almost similar with CMIM. However, it avoid $I(X_j; X_k) - I(X_j; X_k|Y)$ becomes negative. Thus, the score can be calculated by $\jmath_{ICAP}(X_k) = I(X_k; Y) - \sum_{x_j \epsilon S} \max_{x_j \epsilon S}[0, I(X_j; X_k) - I(X_j; X_k|Y)]$ [51].
10	Conditional redundancy (CONDRED)	CONDRED is defined to fulfil unexplored region introduced by the condition likelihood maximization framework ($\beta = 0$) and ($\gamma = 1$) . This criterion that can be expressed by $\jmath_{CONDRED}(X_k) = I(X_k; Y) + \sum_{x_j \epsilon S} I(X_j; X_k|Y)$ [47].
11	Fast correlation-based filter (FCBF)	FCBF has two principles for feature selection including F-correlation and C-correlation. F-correlation denotes the correlation between features. On the other side, C-correlation is correlation between features and class label. They are measured using normalized version of mutual information called symmetrical uncertainty (SU) that can be expressed by $SU(X_S, Y) = 2\frac{I(X_S; Y)}{H(X_S) + H(Y)}$. Then, heuristic mechanisms are utilized on the feature subsets to get the most relevant and to neglect redundant features [52].
12	Fisher score	Fisher score is similarity-based filter FSA. Fisher score for feature X_k can be calculated by $fisher\_score(X_k) = \frac{\sum_{l=1}^{c} n_l(\mu_{kl} - \mu_k)}{\sum_{l=1}^{c} n_l \sigma_{kl}^2}$, where n_l, μ_k, μ_{kl} and σ_{kl}^2 denote the number of sample in class l, mean of feature X_k , mean of feature X_k in class l, variance of feature X_k in class l, respectively. Top m features is obtained by the largest fisher score [53].
13	ReliefF	The basic idea of ReliefF is to assess how well a particular feature distinguishes between instances that are near to each other. Based on randomly selected instances l , ReliefF searches for its two nearest neighbors: one from the same class, called nearest hit H, and the other from the different class, called nearest miss M . The score for quality estimation of feature X_k can be formulated by $ReliefF\_score(X_k) = \frac{1}{c} \sum_{j=1}^{l} (-\frac{1}{h_j} \sum_{x_r \epsilon NH_j} d(x_{jk} - x_{rk}) + \sum_{y \neq y_j} \frac{1}{m_{jy}} \frac{ratio_y}{1 - ratio_y} \sum_{x_r \epsilon NM_{jy}} d(x_{jk} - x_{rk}))$, where c, NH_j, NM_{jy} are the number of classes, the nearest instance of x_j in the same class, and in class y, respectively. $h_j, m_{jy}, ratio_y$ are size of NH_j, size of NM_{jy} , ratio of instances in class y, respectively [49], [54].
14	Trace ratio	The basic idea of trace ratio is to minimize the similarity of instances from different classes and to maximize the similarity of instances from the same class. Trace ratio generates two affinity matrices to identify within-class and between-class data similarity. The trace ratio score of selected m features in S can be formulated by $traceratio(S) = \frac{tr(W'X'L_b XW)}{tr(W'X'L_w XW)}$, where L_b, L_w are Laplacian matrices that correspond to two affinity matrices S_b and S_w, respectively. $tr(.), W$ and X indicate the trace of a matrix, the selection indicator matrix, and the training data, respectively [55].
15	Chi-square	Chi-square feature selection performs independency test to assess whether the feature depends or not depends on the class label. The high chi-square score indicates that a feature is relatively important. Chi-square score can be calculated by $chi\_square(X_k) = \sum_{i=1}^{n} \sum_{j=1}^{c} \frac{(m_{ij} \mu_{ij})^2}{\mu_{ij}}$, where m_{ij} is the number of instances with the ith feature value from feature X_k. Furthermore, $\mu_{ij} = \frac{m_{*i} m_{*j}}{m}$, where m_{*i} is the number of instances with i^{th} feature value from feature X_k. m_{*j} denotes the number of instances in class c [56].
16	F-score	This method implements the ANOVA f-value feature selection. The F-score value can be calculated as follows: $F = \frac{MS_B}{MS_W}$, where MS_B defines as $\frac{\sum_i n_i(\overline{x}_i - \overline{x})^2}{m-1}$ and MS_W defines as $\frac{\sum_{ij} n_i(\overline{x}_{ij} - \overline{x}_i)^2}{n-m}$ that characterizes between and within group variance, respectively. Here, n_i denotes the sample mean in the ith group, x_{ij} denotes the jth observation in the ith group and $\overline{x}$ denotes the overall mean of the data [57], [58].
17	Gini index	Gini index is used as statistical measure to calculate if a feature is capable to separate instance from different classes [49], [59]. Gini index can be formulated by $gini\_index(X_k) = \min_z(p(Z)(1 - \sum_{j=1}^{c} p(C_j|Z)^2) + p(\overline{Z})(1 - \sum_{j=1}^{c} p(C_j|\overline{Z})^2))$, where $Z, \overline{Z}$ are the set of instances that the feature value smaller or equal to the ith feature value and larger than the ith feature value, respectively. $p(.)$ and $p(.|.)$ denote probability and conditional probability, respectively.

D. R. Wijaya, F. Afianti: Stability Assessment of FSAs on Homogeneous Datasets: Study for Sensor Array Optimization Problem
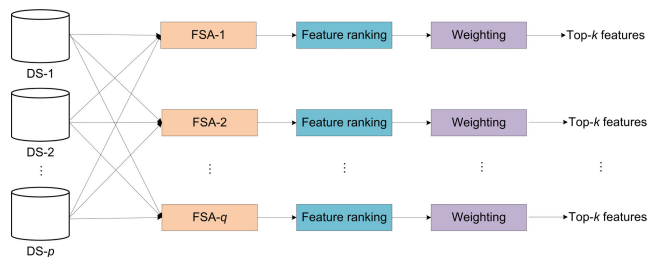
IEEE *Access*

FIGURE 2. Illustration of weighting and feature selection.

as follows:

$$w_{DS_i, x_{jl}} = \frac{|x_j| - rank_{DS_i, x_{jl}} + 1}{|x_j|},\tag{2}$$

where $rank_{DS_i, x_{jl}}$ means rank of feature x_{jl} in dataset DS_i. Sorting a feature output into a top-k features, selected feature set can be determined according to this rule

$$y_i = \begin{cases} 1, & \text{if } w_{y_i} > \overline{w} \\ 0, & \text{if } w_{y_i} \leq \overline{w}, \end{cases}\tag{3}$$

where 0 and 1 indicate the feature is not selected and selected, respectively. w_{y_i} and $\overline{w}$ are weight of feature y_i and weight average of overall features, respectively. Furthermore, Fig. 2 illustrates the flow of experiment where FSAs produce feature ranking from each dataset. Top-k features according to a particular FSA in every dataset are determined by a weighted appearance. Hence, we have several top-k features based on the combination of FSA and datasets. Thus, the stability of FSA is measured according to the difference in output between the datasets (cross-stability).

C. STABILITY METRIC

In works of literature, several stability metrics are proposed for example Jaccard [20] and Hamming [37] distance. However, the drawback is explained that they susceptible to subset-size-biased that implies the consistency in different settings is questionable [60]. On the other hand, to deal with this problem, the stability index was proposed by Kuncheva that is expressed by

$$\mathbb{S}_k(y, \widetilde{y}) = \frac{r - \frac{k^2}{n}}{k - \frac{k^2}{n}} = \frac{rn - k^2}{k(n - k)}\tag{4}$$

where $y, \widetilde{y}, r, k, n$ denote selected feature set in first observation, selected feature set in second observation, number of intersection between selected feature set, number of selected feature set, and total number of features, respectively. This index must satisfy condition $y \subset x$ and $\widetilde{y} \subset x$. Thus, $|y| = |\widetilde{y}| = k$, where $0 < k < |x| = n$. Obviously, several FSAs are possible to produce different cardinality that means this metric cannot be directly used ($k_y \neq k_{\widetilde{y}}$). Moreover, Nogueira proposed the extension of this metric to handle different cardinality in selected feature set [60]. This metric

TABLE 4. Cross-stability format among 12 datasets.

	DS-1	DS-2	...	DS-n
DS-1	$\mathbb{S}_N^{1,1}$	$\mathbb{S}_N^{1,2}$	...	$\mathbb{S}_N^{1,n}$
DS-2	$\mathbb{S}_N^{2,1}$	$\mathbb{S}_N^{2,2}$	...	$\mathbb{S}_N^{2,n}$
...	...	...	...	...
DS-n	$\mathbb{S}_N^{n,1}$	$\mathbb{S}_N^{n,2}$	...	$\mathbb{S}_N^{n,n}$

can be expressed as follows:

$$\begin{aligned}
&\mathbb{S}_N(y, \widetilde{y}) \\
&= \frac{r - \mathbb{E}[r]}{\max[-\max(0, k_y + k_{\widetilde{y}} - n) + \mathbb{E}[r]; \min(k_y + k_{\widetilde{y}}) - \mathbb{E}[r]]} \\
&= \frac{r - \frac{k_y k_{\widetilde{y}}}{n}}{\max[-\max(0, k_y + k_{\widetilde{y}} - n) + \frac{k_y k_{\widetilde{y}}}{n}; \min(k_y, k_{\widetilde{y}}) - \frac{k_y k_{\widetilde{y}}}{n}]}.
\end{aligned}\tag{5}$$

Similar to Kuncheva's index, this measure belongs to $[-1, 1]$. It reaches the maximum value when two selected feature subsets are identical and vice versa. In this experiment, 12 homogeneous datasets are used to investigate the stability of 17 FSAs. For every FSA, the cross-stability of 12 datasets is calculated. Table 4 shows the matrix of stability among 12 datasets, where $\mathbb{S}_N^{a,b}$ means the stability value of ath row and b^{th} column correspond to two different datasets.

IV. STABILITY ASSESSMENT BASED ON THE EXTENSION OF KUNCHEVA'S INDEX

In this experiment, we investigate the stability of 17 FSAs that have been mentioned before. There are three major groups of FSA including information-theoretic (MIM, MRMR, MIFS, CMIM, JMI, DISR, CIFE, ICAP, CONDRED, CMI, FCBF), similarity-based (fisher score, reliefF, trace ratio), and statistical-based (chi-square, F-score, gini index). These outputs can have different cardinality depend on the dataset. It occurs on several FSAs like CMI and FCBF. Hence, the utilization of the extension of Kuncheva's index is reasonable. Fig. 3 shows the heat maps that imply stability cross-dataset on 12 homogeneous datasets. The brighter box color indicates that the selected feature subset on a particular dataset has similar feature members on another dataset. Otherwise, the darker box color means the selected feature subset on a dataset is highly different from another dataset. Thus, a heat map with many dark colors implies an FSA is less stable. According to this figure, MIM, relief, and chi-square get many pale boxes that imply they have the most stable outputs on 12 datasets. However, none of them can produce a consistent feature subset in each dataset.

Moreover, Fig. 4 shows the stability box plot of FSAs. CMIM, JMI, DISR, CIFE, and ICAP show almost similar stability to determine feature subset recommendation. MIM, reliefF, and chi-square produce the best stability for information-theoretic, similarity, and statistical-based FSA, respectively. Furthermore, chi-square becomes the most stable FSA that is denoted by the average value of the
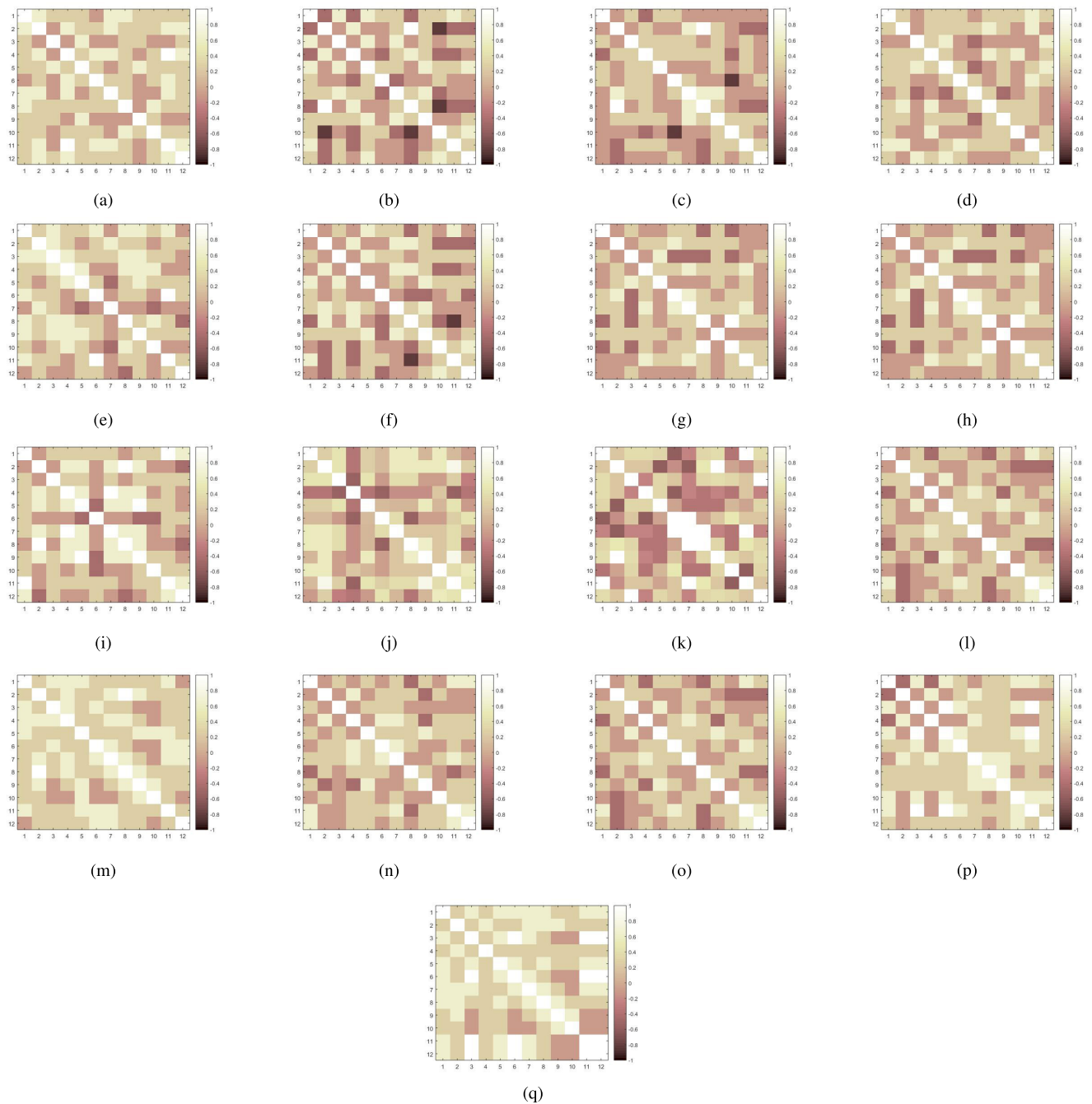
IEEE *Access*

D. R. Wijaya, F. Afianti: Stability Assessment of FSAs on Homogeneous Datasets: Study for Sensor Array Optimization Problem

FIGURE 3. FSA stability heat map on 12 homogeneous datasets: (a) MIM; (b) MRMR; (c) MIFS; (d) CMIM; (e) JMI; (f) DISR; (g) CIFE; (h) ICAP; (i) CONDRED; (j) CMI; (k) FCBF; (l) fisher score; (m) reliefF; (n) trace_ratio; (o) f−score; (p) gini_index; (q) chi−square.

stability index. It is because chi-square has several high outliers ($\mathbb{S}_N = 1$), but has no low outlier. However, none of them can produce consistent output that makes satisfactory generalization of sensor combination. In addition, several FSAs produce negative outlier that implies they have almost completely different outputs on a different dataset. Hence, it is too risky to use a single algorithm to decide which the best sensor combination in the case of sensor array optimization.

Table 5 shows the summary of performance such as computational complexity, redundancy term, and average

stability. It can become consideration for practitioners related to utilization of FSAs. In this study, computational complexity is observed based on the library implementation of each FSA. The existence of redundancy term implies that FSAs have a mechanism to deal with feature redundancy. Moreover, average stability is obtained from cross-stability among 12 homogenous datasets. The majority of information-theoretical FSAs consider both feature relevancy and feature redundancy. According to Table 3, redundancy term is represented by $\sum_{X_j \in S} I(X_k; X_j)$. Based on

D. R. Wijaya, F. Afianti: Stability Assessment of FSAs on Homogeneous Datasets: Study for Sensor Array Optimization Problem
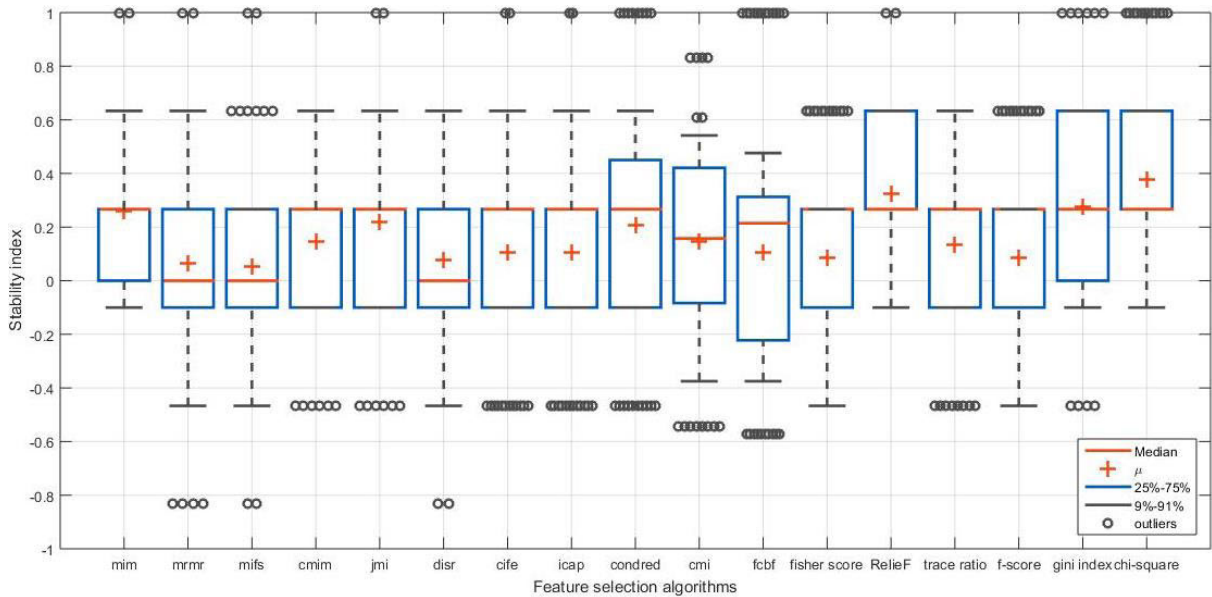
IEEE *Access*

FIGURE 4. FSA stability box plot.

experimental results, FSAs with redundancy term have relatively low stability including MIFS, MRMR, CIFE, CMIM, DISR, ICAP, CMI, FCBF. Furthermore, MIM doesn't have a redundancy term and has relatively low computational complexity ($O(n^2)$). It gets the highest output stability in the information-theoretic FSA group. CONDRED also gets the third highest stability value because it only considers conditional redundancy. On the other side, JMI gets relatively high stability even though it has a redundancy term (0.21898). This result supports the experiment performed by Brown et.al that JMI has the best trade-off as well as balances the relevancy and redundancy terms and includes the conditional redundancy [47]. Different from information-theoretical FSAs, similarity and statistical-based FSAs assess the importance of feature individually without feature redundancy handling. Hence, they only emphasize how to find the most relevance features related to class label. Typically, similarity-based FSA is simple and straight-forward. The focus of computation is to build an affinity matrix for score calculation. Thus, the computational complexity is no more than $O(n^3)$. ReliefF obtains the highest stability (0.32593). Fisher score is less stable because it highly relies on statistical measures such as mean and variance. Their values can vary even though on homogeneous datasets. Furthermore, statistical-based FSAs also individually evaluate the importance of features so they cannot handle feature redundancy. Most of them depend on predefined statistical measures to eliminate irrelevant features. Chi-square has the highest stability even for all groups (0.37685). It uses a simple scoring and has low computational complexity ($O(n^2)$). For F-score, it employs mean value to characterize between and within group variance. As well as Fisher score, it produces very low stability (0.08657). In addition, Gini index gets 0.275 stability score with $O(n^4)$ of computational complexity. The computational complexity

TABLE 5. Summary of FSAs performance.

FSAs	Computational complexity	Redundancy term	Average stability
MIM	$O(n^2)$	No	0.25972
MIFS	$O(n^4)$	Yes	0.05093
MRMR	$O(n^4)$	Yes	0.0662
CIFE	$O(n^4)$	Yes	0.10694
JMI	$O(n^4)$	Yes	0.21898
CMIM	$O(n^4)$	Yes	0.14769
DISR	$O(n^4)$	Yes	0.07639
ICAP	$O(n^4)$	Yes	0.10694
CONDRED	$O(n^4)$	No	0.2088
CMI	$O(n^3)$	Yes	0.14593
FCBF	$O(n^3)$	Yes	0.10655
Fisher score	$O(n^2)$	No	0.08657
RelieF	$O(n^3)$	No	0.32593
Trace ratio	$O(n^3)$	No	0.13241
Chi-square	$O(n^2)$	No	0.37685
F-score	$O(n^2)$	No	0.08657
Gini index	$O(n^4)$	No	0.275

of Gini index is relatively high because it uses probability as a basis for selection criterion. Overall results show that a single FSA cannot present stable outputs that implies cannot guarantee the generalization of sensor array.

V. CONCLUSION
The stability of FSA becomes a major concern in many real-world applications. In this experiment, 12 homogeneous datasets from e-nose were generated. FSA is typically used to deal with sensor array optimization problem. The investigation was performed using 17 filter-based FSAs and cross-stability was measured among datasets. The experimental results show that a single FSA generates different selected features set on the homogeneous dataset. Majority of information-theoretic-based FSAs produce low

IEEE *Access*

D. R. Wijaya, F. Afianti: Stability Assessment of FSAs on Homogeneous Datasets: Study for Sensor Array Optimization Problem

stability. The most stable FSA in this group is MIM that not considers feature redundancy and low computational complexity. It can be applied as a first step to select features before performing a more advanced FSA. JMI is the most stable FSA that considers both relevancy and redundancy. ReliefF and chi-square are the most stable in similarity-based and statistical-based FSA, respectively. They have low computational complexity but individually evaluate feature relevancy without considering redundancy among selected features. According to the stability index, all of FSAs cannot present satisfactory performance that means they have unstable outputs. Unfortunately, in e-nose applications, it becomes a serious problem because different gas sensor combination must be used to classify various beef types. Hence, the sensor array has a lack of generalization. It is not a wise choice to use a different combination of sensors to detect a different sample. Another option is to use as many sensors as possible to detect samples. However, it is also not a good solution because it will raise other problems such as high electrical consumption, high-dimensional data, computational overhead, high data traffic, etc. The main objective of sensor array optimization is to determine the best sensor combination which not only contains the fewest possible number of sensors but also has good generalization. This should be a concern for researchers and practitioners to optimize sensor arrays in the proper way to guarantee generalization. Hence, there is necessary to develop the proper approach to deal with this problem. For future work, the hybrid FSA will be developed to improve the stability of a single FSA.

ACKNOWLEDGMENT

REFERENCES

[1] D. R. Wijaya, R. Sarno, and E. Zulaika, "Gas concentration analysis of resistive gas sensor array," in *Proc. Int. Symp. Electron. Smart Devices (ISESD)*, Nov. 2016, pp. 337–342.

[2] Hariyanto, R. Sarno, and D. R. Wijaya, "Detection of diabetes from gas analysis of human breath using e-Nose," in *Proc. 11th Int. Conf. Inf. Commun. Technol. (ICTS)*, Oct. 2017, pp. 241–246.

[3] D. R. Wijaya, R. Sarno, and A. F. Daiva, "Electronic nose for classifying beef and pork using Naïve Bayes," in *Proc. Int. Seminar Sensors, Instrum., Meas. Metrol. (ISSIMM)*. Surabaya, Indonesia, Aug. 2017, pp. 104–108. [Online]. Available: https://ieeexplore.ieee.org/document/8124272

[4] M. Nurjuliana, Y. B. C. Man, and D. M. Hashim, "Analysis of lard's aroma by an electronic nose for rapid Halal authentication," *J. Amer. Oil Chemists' Soc.*, vol. 88, no. 1, pp. 75–82, 2011.

[5] R. Sarno and D. R. Wijaya, "Recent development in electronic nose data processing for beef quality assessment," *TELKOMNIKA (Telecommun. Comput. Electron. Control)*, vol. 17, no. 1, p. 337, Feb. 2019.

[6] D. R. Wijaya and R. Sarno, "Mobile electronic nose architecture for beef quality detection based on Internet of Things technology," *Bandung*, vol. 2, pp. 655–663, Feb. 2015. [Online]. Available: http://globalilluminators.org/conferences/gtar-2015-indonesia/gtar-full-paper-proceeding-2015/

[7] D. R. Wijaya, R. Sarno, E. Zulaika, and S. I. Sabila, "Development of mobile electronic nose for beef quality monitoring," in *Proc. 4th Inf. Syst. Int. Conf.*, vol. 124, B. V. Bali. Amsterdam, The Netherlands: Elsevier, Nov. 2017, pp. 728–735. [Online]. Available: https://doi.org/10.1016/j.procs.2017.12.211

[8] D. R. Wijaya, R. Sarno, and E. Zulaika, "Information quality ratio as a novel metric for mother wavelet selection," *Chemometric Intell. Lab. Syst.*, vol. 160, pp. 59–71, Jan. 2017, doi: 10.1016/j.chemolab.2016.11.012.

[9] D. R. Wijaya, R. Sarno, and E. Zulaika, "Noise filtering framework for electronic nose signals: An application for beef quality monitoring," *Comput. Electron. Agricult.*, vol. 157, pp. 305–321, Feb. 2019. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0168169918300723

[10] D. R. Wijaya, R. Sarno, and E. Zulaika. (2018). *Electronic Nose Dataset for Beef Quality Monitoring Under an Uncontrolled Environment*. [Online]. Available: https://data.mendeley.com/datasets/mwmhh766fc/2

[11] K. Triyana, M. Taukhid Subekti, P. Aji, S. Nur Hidayat, and A. Rohman, "Development of electronic nose with low-cost dynamic headspace for classifying vegetable oils and animal fats," *Appl. Mech. Mater.*, vol. 771, pp. 50–54, Jul. 2015.

[12] K. Triyana, A. Masthori, B. P. Supardi, M. Iqbal, and A. Bharata, "Prototype of electronic nose based on gas sensors array and back propagation neural network for tea," *J. Math. Natural Sci.*, vol. 17, no. 3, pp. 57–62, 2007.

[13] D. Lelono, K. Triyana, S. Hartati, and J. E. Istiyanto, "Principal component analysis classification of Indonesia black teas based on quality by using electronic nose and principal component analysis," in *Proc. AIP Conf.*, vol. 1755, 2016, Art. no. 020003. [Online]. Available: http://scitation.aip.org/content/aip/proceeding/aipcp/10.1063/1.4958468

[14] S. N. Hidayat and K. Triyana, "Optimized back-propagation combined with radial basic neural network for improving performance of the electronic nose: Case study on the fermentation process of tempeh," in *Proc. AIP Conf.*, vol. 1755, Jul. 2016, Art. no. 020001.

[15] D. R. Wijaya, R. Sarno, and E. Zulaika, "Sensor array optimization for mobile electronic nose: Wavelet transform and filter based feature selection approach," *Int. Rev. Comput. Softw. (IRECOS)*, vol. 11, no. 8, p. 659, Aug. 2016.

[16] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Comput. Appl.*, vol. 24, no. 1, pp. 175–186, Mar. 2013.

[17] Y. Saeys, I. Inza, and P. Larranaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, Aug. 2007.

[18] U. M. Khaire and R. Dhanalakshmi, "Stability of feature selection algorithm: A review," *J. King Saud Univ.-Comput. Inf. Sci.*, to be published, doi: 10.1016/j.jksuci.2019.06.012.

[19] P. Yang, B. B. Zhou, J. Y.-H. Yang, and A. Y. Zomaya, "Stability of feature selection algorithms and ensemble feature selection methods in bioinformatics," in *Biological Knowledge Discovery Handbook*, M. Elloumi and A. Y. Zomaya, Eds. Hoboken, NJ, USA: Wiley, 2013, ch. 14, pp. 333–352.

[20] A. Kalousis, J. Prados, and M. Hilario, "Stability of feature selection algorithms: A study on high-dimensional spaces," *Knowl. Inf. Syst.*, vol. 12, no. 1, pp. 95–116, Dec. 2006.

[21] É. Perthame, C. Friguet, and D. Causeur, "Stability of feature selection in classification issues for high-dimensional correlated data," *Statist. Comput.*, vol. 26, no. 4, pp. 783–796, May 2015.

[22] D. Dernoncourt, B. Hanczar, and J.-D. Zucker, "Analysis of feature selection stability on high dimension and small sample data," *Comput. Statist. Data Anal.*, vol. 71, pp. 681–693, Mar. 2014, doi: 10.1016/j.csda.2013.07.012.

[23] A. Baranov, D. Spirjakin, S. Akbari, and A. Somov, "Optimization of power consumption for gas sensor nodes: A survey," *Sens. Actuators A, Phys.*, vol. 233, pp. 279–289, Sep. 2015, doi: 10.1016/j.sna.2015.07.016.

[24] H. Zhang, J. Wang, X. Tian, H. Yu, and Y. Yu, "Optimization of sensor array and detection of stored duration of wheat by electronic nose," *J. Food Eng.*, vol. 82, no. 4, pp. 403–408, Oct. 2007.

[25] S. Zhang, C. Xie, D. Zeng, H. Li, Y. Liu, and S. Cai, "A sensor array optimization method for electronic noses with sub-arrays," *Sens. Actuators B, Chem.*, vol. 142, no. 1, pp. 243–252, Oct. 2009.

[26] Z. Xu, X. Shi, and S. Lu, "Integrated sensor array optimization with statistical evaluation," *Sens. Actuators B, Chem.*, vol. 149, no. 1, pp. 239–244, Aug. 2010, doi: 10.1016/j.snb.2010.05.038.

[27] Z. Xu and S. Lu, "Multi-objective optimization of sensor array using genetic algorithm," *Sens. Actuators B, Chem.*, vol. 160, no. 1, pp. 278–286, Dec. 2011, doi: 10.1016/j.snb.2011.07.048.

[28] P. M. Szecówka, A. Szczurek, and B. W. Licznerski, "On reliability of neural network sensitivity analysis applied for sensor array optimization," *Sens. Actuators B, Chem.*, vol. 157, no. 1, pp. 298–303, Sep. 2011. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0925400511002620

D. R. Wijaya, F. Afianti: Stability Assessment of FSAs on Homogeneous Datasets: Study for Sensor Array Optimization Problem

IEEE Access

[29] A. K. Bag, B. Tudu, J. Roy, N. Bhattacharyya, and R. Bandyopadhyay, "Optimization of sensor array in electronic nose: A rough set-based approach," *IEEE Sensors J.*, vol. 11, no. 11, pp. 3001–3008, Nov. 2011.

[30] P. Saha, S. Ghorai, B. Tudu, R. Bandyopadhyay, and N. Bhattacharyya, "Optimization of sensor array in electronic nose by combinational feature selection method," in *Proc. 6th Int. Conf. Sens. Technol. (ICST)*, 2012, pp. 341–346.

[31] P. Jia, F. Tian, S. Fan, Q. He, J. Feng, and S. X. Yang, "A novel sensor array and classifier optimization method of electronic nose based on enhanced quantum-behaved particle swarm optimization," *Sensor Rev.*, vol. 34, no. 3, pp. 304–311, Jun. 2014.

[32] L. Zhang, F. Tian, and G. Pei, "A novel sensor selection using pattern recognition in electronic nose," *Measurement*, vol. 54, pp. 31–39, Aug. 2014. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0263224114001559

[33] J. Yan, S. Duan, T. Huang, and L. Wang, "Hybrid feature matrix construction and feature subset optimization-based multi-objective QPSO for electronic nose in wound infection detection," *Sensor Rev.*, vol. 36, no. 1, pp. 23–33, Jan. 2016. [Online]. Available: http://www.emeraldinsight.com/doi/10.1108/SR-01-2015-0011

[34] K. Xu, J. Wang, Z. Wei, F. Deng, Y. Wang, and S. Cheng, "An optimization of the MOS electronic nose sensor array for the detection of chinese pecan quality," *J. Food Eng.*, vol. 203, pp. 25–31, Jun. 2017. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0260877417300298

[35] H. Wu, T. Yue, Z. Xu, and C. Zhang, "Sensor array optimization and discrimination of apple juices according to variety by an electronic nose," *Anal. Methods*, vol. 9, no. 6, pp. 921–928, 2017.

[36] G. Wei, J. Zhao, Z. Yu, Y. Feng, G. Li, and X. Sun, "An effective gas sensor array optimization method based on random Forest*," in *Proc. IEEE SENSORS*, Oct. 2018, pp. 1–4.

[37] M. Ghasemi-Varnamkhasti, A. Mohammad-Razdari, S. H. Yoosefian, Z. Izadi, and G. Rabiei, "Selection of an optimized metal oxide semiconductor sensor (MOS) array for freshness characterization of strawberry in polymer packages using response surface method (RSM)," *Postharvest Biol. Technol.*, vol. 151, pp. 53–60, May 2019.

[38] K. Dunne, P. Cunningham, and F. Azuaje, "Solutions to instability problems with sequential wrapper-based approaches to feature selection," *Mach. Learn.*, to be published. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.11.4109&rep=rep1&type=pdf

[39] S. Nogueira and G. Brown, "Measuring the stability of feature selection," in *Machine Learning and Knowledge Discovery in Databases*, vol. 9852, P. Frasconi, N. Landwehr, G. Manco, and J. Vreeken, Eds. Cham, Switzerland: Springer, 2016, pp. 442–457. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-46227-1_28

[40] H. Wang, T. M. Khoshgoftaar, and N. Seliya, "On the stability of feature selection methods in software quality prediction: An empirical investigation," *Int. J. Softw. Eng. Knowl. Eng.*, vol. 25, no. 09n10, pp. 1467–1490, Mar. 2016.

[41] J. P. Harley and L. M. Prescott, *Microbiology*, 5th ed. New York, NY, USA: McGraw-Hill, 2002, no. 1.

[42] C. Food and N. Sciences. (Nov. 2003). *Vacuum-Packed Meat: Storage Life and Spoilage*. Brisbane. [Online]. Available: https://meatupdate.csiro.au/VPmeat-spoilage-storage.pdf

[43] D. D. Lewis, "Feature selection and feature extraction for text categorization," in *Proc. workshop Speech Natural Lang. (HLT)*, 1992, pp. 212–217.

[44] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Trans. Neural Netw.*, vol. 5, no. 4, pp. 537–550, Jul. 1994.

[45] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[46] D. Lin and X. Tang, "Conditional infomax learning: An integrated framework for feature extraction and fusion," in *Lecture Notes in Computer Science* (Including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 3951. Berlin, Germany: Springer, 2006, pp. 68–82.

[47] G. Brown, A. Pocock, M.-J. Zhao, and M. Lujan, "Conditional likelihood maximisation: A unifying framework for information theoretic feature selection," *J. Mach. Learn. Res.*, vol. 13, pp. 27–66, Jan. 2012. [Online]. Available: http://jmlr.csail.mit.edu/papers/v13/brown12a.html

[48] H. Yang, "Data visualization and feature selection: New algorithms for nongaussian data," in *Proc. Adv. Neural Inf. Process. Syst.*, 1999, pp. 687–693.

[49] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," 2016, *arXiv:1601.07996*. [Online]. Available: http://arxiv.org/abs/1601.07996

[50] P. E. Meyer, C. Schretter, and G. Bontempi, "Information-theoretic feature selection in microarray data using variable complementarity," *IEEE J. Sel.Topics Signal Process.*, vol. 2, no. 3, pp. 261–274, Jun. 2008.

[51] A. Jakulin, "Machine learning based on attribute interactions," Ph.D. dissertation, Dept. Fac. Comput. Inf. Sci., Univ. Ljubljana, Ljubljana, Slovenia, 2005.

[52] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Int. Conf. Mach. Learn. (ICML)*, 2003, pp. 1–8. [Online]. Available: http://www.aaai.org/Papers/ICML/2003/ICML03–111.pdf

[53] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Hoboken, NJ, USA: Wiley, 2012.

[54] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, nos. 1–2, pp. 23–69, Oct. 2003.

[55] F. Nie, S. Xiang, Y. Jia, C. Zhang, and S. Yan, "Trace ratio criterion for feature selection," in *Proc. Conf. Artif. Intell. (AAAI)*, vol. 2, 2008, pp. 671–676.

[56] H. Liu and R. Setiono, "Chi2: Feature selection and discretization of numeric attributess," in *Proc. 7th IEEE Int. Conf. Tools Artif. Intell.*, Nov. 1995, pp. 388–391.

[57] S. Wright, "The interpretation of population structure by F-Statistics with special regard to systems of mating," *Evolution*, vol. 19, no. 3, p. 395, Sep. 1965.

[58] M. Wielgosz, P. Szczepka, P. Russek, E. Jamro, K. Wiatr, M. Pietroń, and D. Żurek, "Evaluation and implementation of n-Gram-Based algorithm for fast text comparison," *Comput. Informat.*, vol. 36, no. 4, pp. 887–907, 2017.

[59] C. W. Gini, "Variability and mutability, contribution to the study of statistical distributions and relations," *J. Amer. Stat. Assoc.*, vol. 66, pp. 534–544, Mar. 1971.

[60] S. Nogueira and G. Brown, "Measuring the stability of feature selection with applications to ensemble methods," in *Proc. Int. Workshop Multiple Classifier Syst.* Cham, Switzerland: Springer, 2015, pp. 135–146.

DEDY RAHMAN WIJAYA was born in Tulungagung, Indonesia, in 1984. He received the bachelor's degree in informatics engineering from STT Telkom Bandung, Indonesia, in 2006, the Master of Engineering degree from the School of Electrical Engineering and Informatics, Institut Teknologi Bandung, in 2010, and the Ph.D. degree in computer science from the Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia, in 2019. He was a Visiting Researcher with the Pulse Lab Jakarta–United Nations Global Pulse, in 2018, where he worked with machine learning and artificial intelligence. He is currently a Lecturer and a Researcher with the School of Applied Science, Telkom University, Bandung. His research interests include cyber-physical systems, intelligent systems, information, signal processing, and machine learning including their applications.

FARAH AFIANTI was born in Surabaya, Indonesia, in 1986. She received the bachelor's and master's degrees in informatics engineering from Telkom University, Bandung, Indonesia, in 2008 and 2015, respectively, and the Ph.D. degree in electrical engineering from the Institut Teknologi Sepuluh Nopember, Surabaya, in 2019. She is currently a Lecturer with the School of Computing, Telkom University, Bandung. Her research interests are in the areas of wireless communication, cryptography, information security, and sensor networks including their application.