Received: January 16, 2021. Revised: March 9, 2021.

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Ensemble Learning for Optimizing Classification of Pork Adulteration in Beef Based on Electronic Nose Dataset

Malikhah Malikhah[1]* **Riyanarto Sarno[1]** **Shoffi Izza Sabilla[1]**

[1]*Informatics Department, Faculty of Intelligent Electrical and Informatics Technology, Institut Teknologi Sepuluh Nopember Sukolilo, Surabaya, 60111, Indonesia*
* Corresponding author's Email: meli2511@gmail.com

Abstract: Pork and beef are the main resources of red meat in the world. However, not everyone can eat pork because of their religious background or other reasons. Therefore, it is very important to ensure the purity of the meat prior to being consumed. This research applied ensemble learning to optimize the classification on Electronic Nose Dataset for Pork Adulteration in Beef. Ensemble learning is one of a method that is widely used and the most successful method to improve performance. This research used several traditional machine learning algorithms and chose a machine learning algorithm, which produced the best result as the base classifier for ensemble learning to optimize the classification on Electronic Nose Dataset for Pork Adulteration in Beef. There were three ensemble learnings used in this research, namely hard voting, stacking, and bagging. The steps conducted in this research comprise (1) pre-processing by de-noising of the raw signals, (2) statistical feature extraction, (3) feature selection, (4) classification, (5) using ensemble learning to improve the performance, and (6) performance evaluation. The experiment result shows that hard voting ensemble learning using K-nearest neighbors (KNN) as base classifier is able to distinguish well between beef, pork, and pork adulteration in beef to seven classes which obtained 98.33% accuracy.

Keywords: Electronic nose, Ensemble learning, Hard voting, K-nearest neighbors, Pork adulteration.

1. Introduction

Pork and beef are the most common main resources of red meat in the world. These two types of red meat are widely consumed animal-based protein [1]; however, not everyone can eat pork because of their religious background or other reasons. Therefore, it is important to ensure the quality and purity of the meat prior to being consumed.

An electronic nose dataset for pork adulteration in beef used a device called electronic nose to retrieve data regarding odor from several weight combinations of pork and beef [2]. An electronic nose is one of the most powerful tools in the food quality industry because it is low-cost, easy to implement, faster process, and flexible [3-4].

This research aims to improve the classification of the electronic nose dataset for pork adulteration in beef by using ensemble learning. Other traditional machine learning algorithms have been commonly used to perform classifications and predictions. This research used a traditional machine learning that has been shown to produce good result in classification as a base learner in an ensemble learner to optimize the accuracy of the classifier. The use of ensemble learning aims to improve the classification's performance without changing the data. Several studies have used ensemble learning to improve prediction accuracy, such as improve data stream accuracy [5], improve classification techniques on majority voting [6], and dealing with imbalanced data [7].

Several researches related to the classification of different types of meat can be seen in Table 1, the best accuracy of 98.10% is obtained using optimized Support Vector Machine (SVM) with the same dataset that we used in this research. From Table 1, we can see that none of the previous research have implemented ensemble learning. Meanwhile, the use of ensemble learning for other electronic nose application has been used, i.e., electronic nose for

Table 1. Recent research of different type of meat using e-nose

Year	Samples	Methods	Accuracy	Ref
2019	Duck adulteration in mutton	Linear Discriminant Analysis (LDA), Partial least square (PLS), Fish Linear discriminant analysis (FLDA)	98.20%	[8]
2019	Pork and miced beef	LDA, PLS	95%	[9]
2019	Beef, pork and chicken	Pearson Correlation, Support Vector Machine (SVM)	92%	[10]
2019	Pork, beef, and mutton	Discriminant factor analysis (DFA), PCA	94.70%	[11]
2019	Chicken, meat cuts of beef, pork,	PCA, Deep learning	96.90%	[4]
2020	Beef Adulterated with Pork	Extreme learning machine (ELM)	91.27%	[12]
2020	Adulteration of minced beef with duck meat and pork	Discriminant Analysis (DA)	95.80%	[13]
2020	Pork Adulteration in Beef	Optimized SVM	98.10%	[14]

wine properties detection used XG-Boost [15], electronic nose signals for classification of diabetes disease used majority voting with Support Vector Machine (SVM) as their base classifiers [16], Transient Feature Fusion in Electronic Nose used majority voting with Multi-Layer Perceptron (MLP), K-nearest neighbor (KNN), and SVM as their base classifiers [17], and identification of ginsengs using electronic nose used AdaBoost [18].

Previous research related to electronic nose have also attempted to classify multiclass, i.e., a multi-classification for the sweetness of pineapple aroma produced an accuracy of 82% [19], a classification for six classes of civet and non-civet coffee produced and accuracy of 95% [20], and classification for three classes of meats produced an accuracy of 92% [21].

The contributions of this research are explained as follows: (i) the selection of the best number of features for classification and (ii) the selection of the best ensemble learning with the chosen classifier as the base classifiers for optimizing the classification of pork adulteration in beef based on electronic nose dataset.

The rest of this paper is arranged as follows: A detailed explanation of the proposed method is explained in Section 2. The result of experiments and discussions are explained in Section 3. Furthermore, Section 4 explains the detailed evaluation of the proposed method using the confusion matrix and the several metrics obtained. Finally, Section 5 contains the conclusion of the experiments.

2. Proposed methods

This research aims to optimize the performance of the classification with ensemble learning for the electronic nose dataset for pork adulteration in beef. The scheme of the proposed method can be seen in Fig. 1. The dataset contains raw signals obtained from the electronic nose device. The raw data in the form

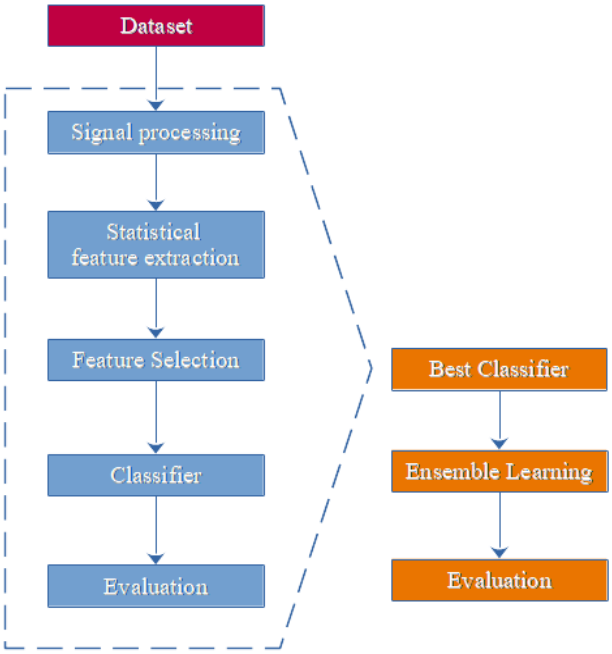


Figure. 1 Proposed method

of digital signals will undergo pre-processing step called signal de-noising to enhance the quality of data input. After signal de-noising, the next step is to extract the statistical parameter to obtain the features. Feature selection aims to choose the best feature subset. Several classifiers were compared to obtain the predictive result, the classifier with the best result would be chosen as the base classifier in ensemble learning. There were three ensemble learning used in the experiment, namely hard voting, stacking, and bagging.

2.1 Dataset

This research used an electronic nose dataset for pork adulteration in beef [2]. The dataset provides seven combinations of weight from the adulteration of pork in beef using the electronic nose device. There were seven classes in this dataset based on

seven combinations of pork and beef mixtures as shown in Table 2.

The data were obtained from the electronic nose device for 120 seconds. The electronic nose device used eight sensors from the MQ family, namely MQ135, MQ137, MQ136, MQ138, MQ9, MQ6, MQ4, MQ2, and two additional sensors, temperature, and humidity. There were 60 samples for each class, so the total of all samples was 420 samples for seven classes.

2.2 Signal processing

Signal processing is a process to enhance the quality of data input in the form of signals. The signal processing was done using discrete wavelet transform (DWT) [22-23]. DWT decomposed the stationary signals into 2D, time, and frequency domains. The wavelet transformation of a signal *x* is calculated using Eq. (1) [24].

$$wt(s,\varphi) = \frac{1}{\sqrt{s}}\int_{-\infty}^{\infty} x(t) w * \left(\frac{t-\varphi}{s}\right) dt \quad (1)$$

where *s* is a parametric scale from the base wavelet $\frac{t-\zeta}{s}$, and $\omega * (.)$ denotes the mother wavelet's conjugation. The φ symbol is a value that changes the wavelet based on the movement of the time axis. The ω is the symbol of mother wavelet. For translation and scaling, DWT used a discrete value as shown in Eq. (2).

$$dwt(i,k) = \left(x(t), \omega_{i,k}(t)\right)$$
$$dwt(i,k) == \frac{1}{\sqrt{2^i}}\int_{-\infty}^{\infty} x(t)\omega * \left(\frac{t-k2^i}{2^i}\right) dt \quad (2)$$

There are several types of mother wavelets and their family wavelets as seen in Table 3. Eq. (3) below is a formula of the signal containing noise.

$$x(t) = y(t) + \theta * n(t) \quad (3)$$

x(t), y(t), n(t), and θ are signals that contain noise, the reconstructed signal, noise, and the level of noise. The main goal of noise filtering is to reduce the noise and maintain the value of reconstructed signals.

2.3 Statistical feature extraction

From the denoised signals produced in signal processing, we calculated or extracted four statistical parameters, namely skewness, kurtosis, standard deviation, and mean value for each signal as features.

2.3.1. Skewness

Table 3. List of mother wavelets and its family

Mother wavelet	Family wavelet
haar	haar
db	From db1 to db10
sym	From sym1 to sym8
coif	From coif1 to coif5
bior	From bior 1.1 to bior 2.8

Table 2. List of classes in the dataset

Label (alternate label name)	Beef (Percentage)	Pork (Percentage)	Number of samples
Class 0 (000)	100	0	60
Class 1 (010)	90	10	60
Class 2 (025)	75	25	60
Class 3 (050)	50	50	60
Class 4 (075)	25	75	60
Class 5 (090)	10	90	60
Class 6 (100)	0	100	60

Skewness is an asymmetrical measure in normal distribution [25]. Skewness can be positive, negative, or zero. Positive skewness means that the tail of the distribution is the right of the most values, indicating that most of the distribution has low values as opposed to negative skewness. Meanwhile, skewness that has zero values means that the value is symmetrically distributed, with the distance between the right and left tail distributions being equal. Eq. (4) is the formula for skewness; where σ is the standard deviation of samples, r_t is the value of the sample observed, μ is the mean of the samples, and *T* is the number of samples observed.

$$S = \frac{1}{T\sigma^3}\sum_{t=1}^{T}(r_t - \mu)^3 \quad (4)$$

2.3.2. Kurtosis

Kurtosis is an indicator that shows the degree of tailedness. The greater the kurtosis value, the taper the curve. Kurtosis can be expressed in Eq. (5); where σ is the standard deviation of samples, r_t is the value of the sample observed, μ is the mean of the samples, and *T* is the amount of samples observed.

$$K = \frac{1}{T\sigma^4}\sum_{t=1}^{T}(r_t - \mu)^4 \quad (5)$$

2.3.3. Standard deviation

Standard deviation is a value used to determine how the distribution is spread out from the mean. Eq. (6) is the formula of standard deviation; where r_t is the value of the sample observed, μ is the mean of the samples, and T is the number of samples observed.

$$\sigma = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(r_t - \mu)^2} \quad (6)$$

2.3.4. Mean

Mean is the average value of the sample observed. Eq. (7) below is used to calculate the mean value; where r_t is the value of the sample observed, μ is the mean of the samples, and T is the number of samples observed.

$$\mu = \frac{1}{T}\sum_{t=1}^{T} r_t \quad (7)$$

2.4 Feature Selection

Feature selection is a method of selecting attributes or features from data that are most relevant to classification. The goal of feature selection is to improve the accuracy of the model by selecting the features that will produce a good result with fewer data. It is expected that with fewer features, the complexity of the model will decrease, and the model becomes simpler. This research used 3 feature selection methods, namely ANOVA F-test, mutual information, and a combination of statistical parameter feature, then compared the results of those three methods.

F-statistics or F-test is part of a statistical test that calculates the ratio between the variance values [26]. ANOVA is a type of F-statistics used when the variables are numerical input, and the target is a categorical type such as classification. The result of the test can be used to select features. If a feature has a high ANOVA f-test value then that feature is significant. ANOVA F-test can be calculated using Eq. (8); where σ_{bcv}^2 is between class variance and σ_{wcv}^2 within-class variance.

$$f - test = \frac{\sigma_{bcv}^2}{\sigma_{wcv}^2} \quad (8)$$

To get the between-class variance value we can use Eq. (9); where k is the number of class, n_i is measurement number, $\bar{X}$ is the mean from all data, and $\bar{X}_i$ is the mean of class i.

$$\sigma_{bcv}^2 = \sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2 \quad (9)$$

Meanwhile, Eq. (10) is used to get within-class variance; where k is the number of class, n_i is measurement number, $\bar{X}$ is the mean from all data, and X_{ij} is the ith measurement of class j.

$$\sigma_{wcv}^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij} - \bar{X}\right)^2 \quad (10)$$

Mutual information is an example of supervised feature selection. Mutual information indicates the amount of information that contributes for making classification decisions. Mutual information works when two independent variables *a* and *b*, which has probability *P(a)* and *P(b)*, where *P(a,b)* is a probability for *a* and *b*.

2.5 Classification methods

This research compared five classification methods to classify seven classes (multiclass) of data in the dataset. Those five classification methods were K-nearest neighbor (KNN), Logistic regression (LR), SVM, LDA, and Naive Bayes (NB). The classification method or classifier which produced the best result was chosen as the base classifier of ensemble learning. To compare several classification methods, k-fold cross-validation is used.

KNN is a type of classification based on distance measurement [27]. KNN assumes that the entire training set not only consists of the data in the set but also includes the classification results of the data points; in other words, the training data becomes a model. In KNN, we first determine the *k* parameter that refers to the number of nearest neighbors, then when there is new data to be classified, the model calculates the new data distance with *k*-nearest data in the training set. There are three ways to measure the distance metric, namely Minkowski, Manhattan, and Euclidean. The most common way to measure the distance metric is Euclidean distance, it can be calculated using Eq. (11); where $d(x, y)$ is the euclidean distance of data points *x* and *y*.

$$d(x, y) = \sqrt{\sum_{j=1}^{k}(x_j - y_j)^2} \quad (11)$$

After calculating the distance metric, the new data will then be placed in the class containing the most data from the *K*-nearest data. KNN can also be used for multiclass classification problems.

LR is a classification method used to estimate discrete values, for example, binary values like 0 and 1, yes and no, or true and false, based on independent variables [28]. LR predicts the occurrence of a value by adjusting the data into the sigmoid function.

Because LR predicts a probability, the output or target value is between 0 and 1. However, LR can also be used to perform multiclass classification, by using the SoftMax function instead of the sigmoid function.

SVM is a classification method that plots each data item as a data point in n-dimensional space, where n is the number of features and the feature value as its coordinate [29]. SVM will draw a line or a plane that will separate the data into two different classes. The same principle is used to solve multiclass problems.

NB is a classification method based on the Bayes' theorem [30]. NB assumes that the existence of a feature in a class is not related to the existence of other features. For multiclass problems, NB calculates the probability for each class then the class with the largest probability will become the output class.

LDA is one of the methods in statistics, it is a pattern recognition to find linear combinations of features that characterize or separate two or more classes [31]. LDA makes predictions by estimating the likelihood that a set of inputs will fit into each class. The class that has the highest probability is the output class. LDA uses the Bayes theorem to estimate the probability of class *k* output with input *x* using the probability of each class and the probability of the data in each class.

2.6 Ensemble Learning

Unlike ordinary single learning, ensemble learning builds a model from one or more classifiers. [32]. An ensemble learning consists of several classifiers called base classifiers. The base classifier can be homogenous or heterogeneous. Homogeneous ensemble learning consists of a single-type base classifier, while heterogeneous ensemble learning consists of different types of base classifiers. An ensemble learning architecture can be seen in Fig. 2. This research built 3 types of ensemble learning, namely majority voting, bagging, and stacking, which use the best classifier obtained from the previous stage as the base classifier so the proposed ensemble learning is constructed from a single type of base classifier or homogeneous ensemble learning.

Pseudo-code for the proposed ensemble learning can be seen in Fig. 3. The input for the proposed ensemble learning is (i) *N*, which is the number of ensemble learning to be compared (bagging, stacking, and majority voting), (ii) *D*, which is pre-processed dataset from the previous stage, (iii) F^D, which is the F-test value of each feature in Dataset D, (iv) *k*, which is the number of selected features, and (v) *T*, which is

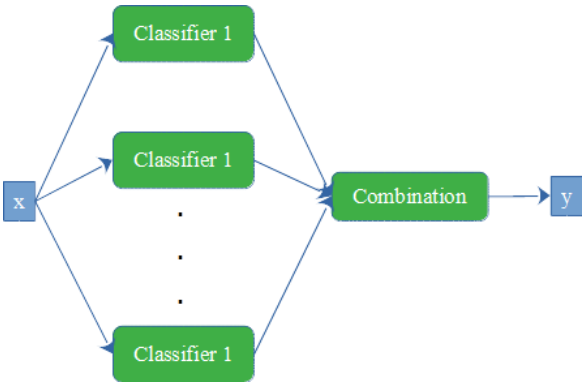


Figure. 3 An ensemble learning architecture

Input : *N* as number of ensemble learning method
D as Pre-processed dataset
F^D as *F-test* value of each feature in D
k as number of selected features
T as base classifier or base model
Ouput : *Best* as chosen ensemble learning method
Step 1 : build ensemble learning *E*
for *n* = 1 to *N* **do**
$(Acc, E)_n = E[n, (D, F^D, k), T]$
end
Step 2 : Chose the best *E*
Best = *max*(*Acc*, *E*)
if *Best* > *Acc*(*T*)
return *Best*
end

Figure. 2 Algorithm of the Proposed ensemble learning

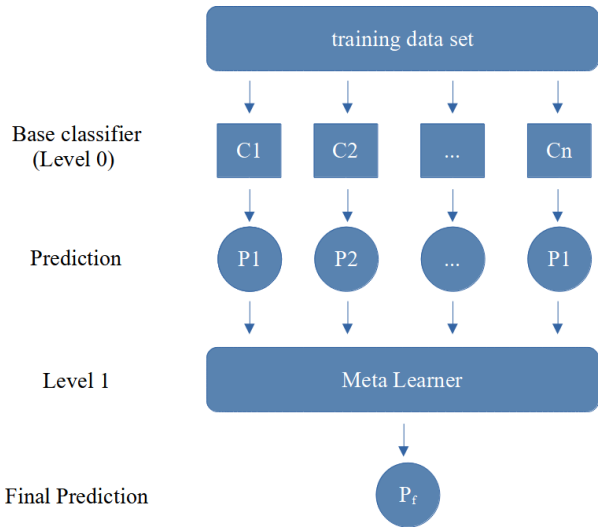


Figure. 4 The architecture of stacking

the base classifier obtained from step 2.5. The first step is building each ensemble learning method (*E*) and calculate its accuracy (*Acc*). Each ensemble learning method E_n will have *T* as base classifiers and trained with the pre-processed dataset *D* using *k* selected features from F^D, then we calculate its accuracy. The second step is to choose the best *E*. The ensemble learning method *E* which produced the

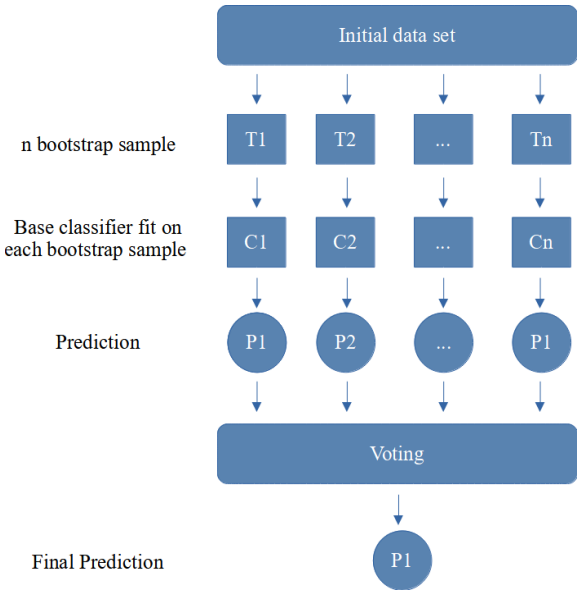


Figure. 5 The architecture of bagging

highest accuracy and can improve the accuracy of the base classifier T will be chosen.

2.6.1. Bagging

Bagging comes from two words, bootstrap and aggregating. Bagging uses bootstrap distribution sampling to retrieve the data subsets for the training process of the base classifier. The idea of bootstrap sample is, if there is a training dataset consisting of n number training examples then a subset of n training examples will be produced by sampling with replacement. By performing the process x times, x subsets sample from n training examples are retrieved. To determine the class of data, bagging uses voting to aggregate the predictions of base classifiers. The architecture of bagging can be seen in Fig. 4. In Fig. 4, T_n are the sample dataset, C_n are the base classifier, and P is the prediction result. Bagging reduces variance and helps avoid overfitting. The final prediction of bagging can be calculated using Eq. 12, where $H(x)$ is the final prediction of bagging; T is the number base classifier; $h_t(x)$ is the prediction for each bootstrap sample; and $\arg max$ is used to aggregate the prediction of base classifiers with voting.

$$H(x) = \underset{y \in Y}{\arg max} \sum_{t=1}^{T} h_t(x) \quad (12)$$

2.6.2. Majority Voting

Each base classifier votes for one class in majority voting. The final prediction class is a class that obtains more than half of the votes. That type of majority voting is called hard voting. Another type of majority voting is soft voting, the probability vector for each predicted class for every classifier are summed up and averaged. The final prediction class label is the one corresponding to the highest value. This research used hard voting as one of the ensemble learning. Fig. 5 is the architecture of hard voting. In Fig. 5, C_n are the base classifier and P is the prediction result. The final prediction of majority voting can be calculated using Eq. 13, where $H(x)$ is the final prediction of majority voting; T is the number base classifiers; $h_i^y(x)$ is the prediction for each base classifier; and $\arg max$ is used to choose the class which has the most votes.

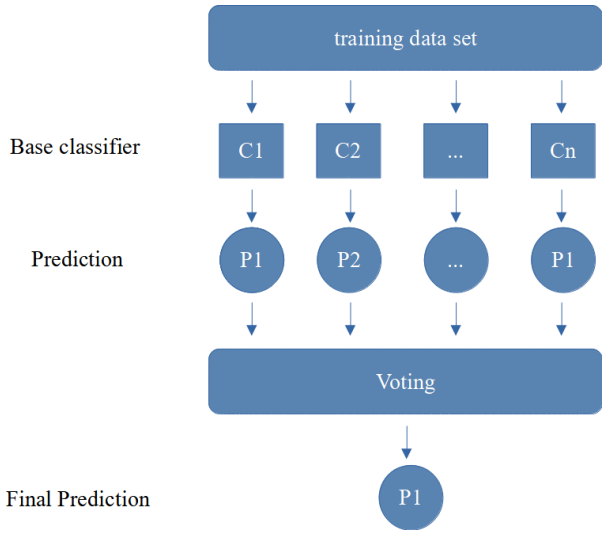
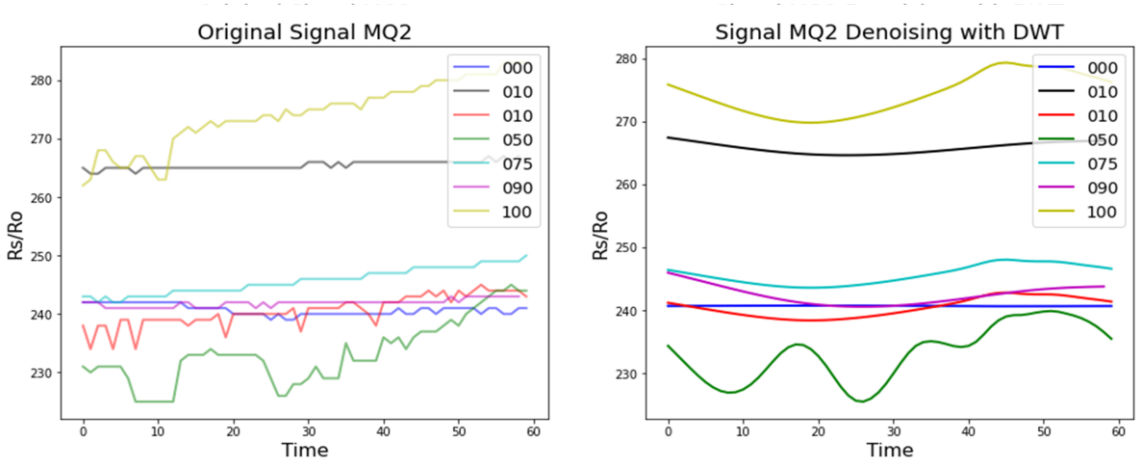


Figure. 6 The architecture of majority voting

$$H(x) = \underset{y \in Y}{\arg max} \sum_{i=1}^{T} h_i^y(x) \quad (13)$$

2.6.3. Stacking

Stacking is an ensemble learning where a classifier is trained to combine several base classifiers. There are two types of learners in stacking, namely first-level learners and meta-learner. First-level learner is the base classifier and meta-learner is a classifier that combines first-level learner. The architecture of stacking can be seen in Fig. 6. In Fig. 6, Cn is the base classifier and P is the prediction result. There are two main steps in stacking; the first step is every base classifier in the training phase uses the same data set and produces each prediction result. The next step is the meta-learner takes the prediction results produced by the base classifier in the first step as input features. Stacking ensemble is often heterogeneous because it uses different classifier algorithms as base classifiers. The final prediction of stacking is calculated using Eq. 14, where $H(x)$ is the final prediction of stacking; T is the number base



(a) (b)

Figure. 7 Response signals: (a) Raw signal of MQ2 (b) Reconstructed signal of MQ2 after signal de-noising

classifiers; h' is the prediction of meta-learner; and $h_T(x)$ is the prediction of first-level learners.

$$H(x) = h'(h_1(x), \dots, h_T(x)) \quad (14)$$

3. Results and discussion

3.1 Result of signal processing

The electronic nose dataset for pork adulteration in beef in the form of signals required signal pre-processing to clear high noise interference. DWT was used to reduce the noise on raw signals from the dataset. This research used combinations of mother wavelets, i.e. haar, db, sym, coif, bior, and rbior also level decomposition from 1 to 11. To determine the best combination of those two parameters, 1650 experiments were conducted. The best combination which produced the best classification result would be chosen as the level of decomposition and the mother wavelet to reduce the noise of raw signals.

From the experiments as seen in Table 4, the best result was obtained from mother wavelet db9 and level of decomposition 5. Meanwhile in Fig. 7 (a) we can see the comparison between one of raw signal (sensor MQ2) with different combinations of weight from the adulteration of pork in beef and Fig. 7 (b) shows the de-noised signal after DWT signal pre-processing with mother wavelet db9 and level of decomposition 5; the de-noised signals indicates smoother result than the raw signals.

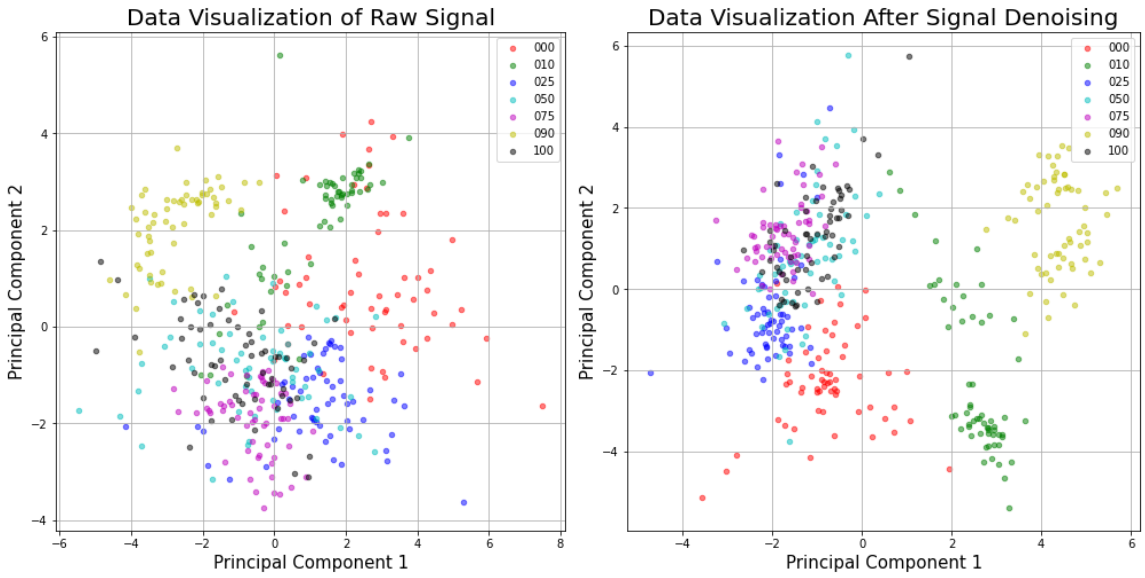
PCA is a technique used to reduce the dimensionality of data but still able to maintain the characteristics of the data. With smaller dimensions, the data can be more easily visualized and analyzed. This research used PCA to visualize the data and then compare the data before and after signal pre-processing. Fig. 8 shows data visualization with PCA. It can be concluded that after signal processing, the distance between classes has become clearer.

Table 4. Combination of mother wavelet and level of decomposition

Number of experiments	Wavelet	Level	Classifier	Accuracy
1	haar	1	LR	91.42%
2	haar	2	LDA	93.33%
3	haar	3	KNN	95.48%
4	haar	4	NB	81.67%
5	haar	5	SVM	61.19%
	...	...	...	...
516	db9	5	LR	91.67%
517	db9	5	LDA	93.33%
518	**db9**	**5**	**KNN**	**96.42%**
	...	...	...	...
1649	bior2.8	11	NB	81.90%
1650	bior2.8	11	SVM	61.19%

3.2 Result of feature extraction of statistical parameters

The electronic nose dataset for pork adulteration in beef has 420 samples, but each sample has 60 rows of frequency signals obtained from an electronic nose device. This research calculated four statistical parameters as mentioned in Section 2 to extract the features of each sample. The statistical parameter features extracted are 40 features, since each sample contains 10 sensors. In Table 5, we can see the classification result of the raw signals prior to signal



(a) (b)

Figure. 9 Data visualization with PCA (a) Raw Signal (b) After signal denoising

Table 5. Classification result

Classifier	Accuracy of raw signals	Accuracy of de-noised signals
LR	92.14%	92.14%
LDA	93.10%	93.33%
KNN	**95.71%**	**96.43%**
NB	81.90%	80.24%
SVM	61.19%	60.71%

pre-processing. KNN produces the best classification result with 95.71%, followed by LDA with 93.10%. Upon the completion of signal pre-processing with DWT, KNN remains producing the best result and improved the accuracy from 95,71% to 96,43%, followed by LDA with an accuracy of 93.30% which also improved from the previous result of 93.10%. From this result, we can conclude that signal processing helps to improve the classification result because it reduces the noise in the raw signals and enhances the quality of data input.

3.3 Results of feature selection

This research used three feature selection methods, namely ANOVA F-value, mutual information, and a combination of statistical parameter features, then compared the results of those three methods. Feature selection which produced the best result and fewest features would be chosen. From the previous result, KNN is chosen as the base classifier because it produced the best accuracy. Table 6 shows that ANOVA with 35 features produced the same classification accuracy as 40 features (all features). The features deleted were MQ2_standard_deviation, MQ138_skewness, MQ2_kurtosis, MQ6_kurtosis, and temp_kurtosis.

Table 6. Feature selection result comparison

Feature selection	Number of features	Accuracy with KNN
ANOVA	**35 features**	**96.43%**
Mutual Information	39 features	96.43%
Mean-standard deviation	20 features	96.19%
skewness-kurtosis	20 features	41.19%
All features	40 features	96.43%

3.4 Classification result

Cross-validation K-fold was used to evaluate the performance of classifiers. It was used because it can reduce the computation time while maintaining the accuracy of the result. This research used K-fold with 10 folds, where the data was divided into 10 folds of the same size. For each subset, cross-validation used nine-folds for training and the remaining fold for testing. There were aforementioned five classifiers compared, namely LDA, LR, KNN, NB, and SVM. The seven classes that classified from dataset A can be seen in Table 2.

Metrics accuracy was used to measure the classification result. The accuracy value shows how accurately each classifier performed the classification. Eq. (15) is the formula used to calculate the accuracy value, where TP, TN, FP, and FN are True Positive, True Negative, False Positive, and False negative respectively.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\% \quad (15)$$

Based on Table 5, KNN is the classifier that produces the best result with an accuracy of 96.43%; therefore KNN was chosen as the base classifier of the ensemble learning in the next experiment.

3.5 Ensemble learning result

The ensemble learning algorithm used in this experiment is hard voting, bagging, and stacking, in which each algorithm uses KNN as the base classifier. Fig. 9 shows that hard voting can improve the accuracy of KNN to 98.33% compared to other ensemble learning algorithms. Hard voting can improve the accuracy because this experiment uses a single-type base classifier (KNN) which already has good performance and all base classifiers in this ensemble mostly agree on the predictive task which is good because hard voting predicts the class label with the most votes.

Bagging produces a lower result with an accuracy of 95.24% because KNN is a stable classifier, which is less sensitive to perturbation on training data. Based on bagging theory, bagging helps decrease the variance and increased the robustness of the classifier. Using a stable classifier is less advantageous because bagging ensemble will not help improve generalization performance or decrease the variance. In bagging classifier, we need to use an unstable classifier like Decision Tree or Random Forest, so it can help reduce variance and eventually improve the accuracy. Stacking also produces a lower result than KNN with an accuracy of 85.48% and does not improve accuracy because the base classifier used were not diverse or heterogeneous, so different assumption about how to solve the predictive modelling task was not made. In stacking, the base classifier needs to be heterogeneous, such as combining linear model with Decision Tree, SVM, Neural Network, and other ensemble algorithms.

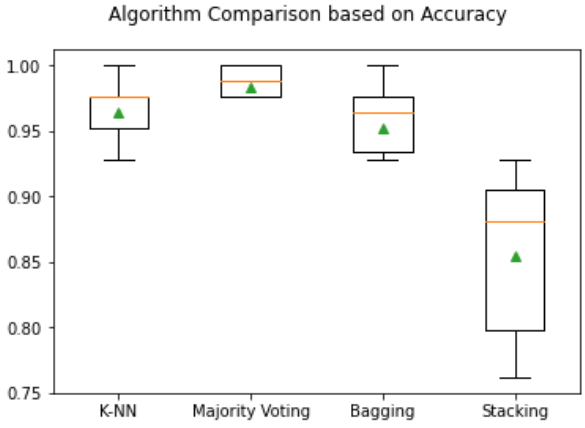


Figure. 10 Ensemble learning accuracy comparison

4. Evaluation and performance

Confusion matrix is a method usually used to measure classifiers performance. Confusion matrix provides information about the comparison of classification results performed by a classification method with the actual class label. In hard voting confusion matrix as seen in Fig. 10, class 000 (class

Table 7. Ensemble learning (hard voting) performance result

Class	Precision	Recall	F1-score	Accuracy
000	100%	100%	100%	100%
010	100%	100%	100%	100%
025	95.16%	98.33%	96.72%	99.05%
050	100%	90%	94.74%	98.57%
075	98.36%	100%	99.17%	99.76%
090	100%	100%	100%	100%
100	95.24%	100%	97.56%	99.29%

Table 8. Comparative study of proposed method

Author	Methodology	Accuracy
Leng [13]	DA	93.33%
Sarno [14]	Optimized SVM	98.10%
Proposed methodology	Ensemble learning (majority voting)	98.33%

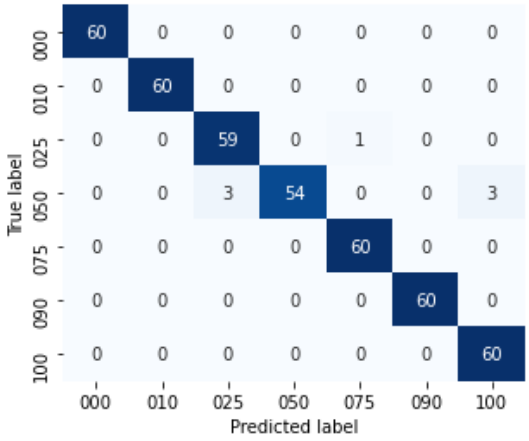


Figure. 11 Confusion matrix of proposed ensemble learning

0), class010 (class 1), class 075 (class 4), class 090 (class 5), and class 100 (class 6) were predicted perfectly. On the other hand, class 025 (class 2) had 1 datum that was wrongly predicted and class 050 (class 3) had 6 data that were wrongly predicted, 3 data were predicted as class 025 and another 3 data were predicted as class 100.

From confusion matrix, other metrics namely precision, recall, and f1 score can be retrieved. The precision value is the ratio of TP predictions to overall positive predicted outcomes. The recall value is a ratio of TP predictions to the overall true positive data, whereas f1 score is a measure of the classifier's accuracy. The precision, recall, f1-score, and accuracy value can be seen in Table 7. Those metrics values show that majority voting produces superior performance in the classification of seven classes from the electronic nose dataset for pork adulteration in beef.

This research also compared the proposed method with the latest previous research using the same dataset in terms of accuracy. The comparison result can be seen in Table 8. It indicates that the proposed method has achieved the best performance with an accuracy of 98.33% compared to DA [13] which produced an accuracy of 93.33% and optimized SVM [14] which produced an accuracy of 98.10%.

5. Conclusion

From the result of the research, we can conclude that signal processing with Discrete Wavelet Transform can affect the result of the classifier accuracy. Raw signals that have undergone signal processing would make the signals smoother and the quality of the signals improved. Feature selection with ANOVA can reduce the number of features to 35 features but can still maintaining the accuracy score result. After comparing five different classifiers, KNN with an accuracy of 96.43% is chosen as the base classifier of the ensemble learning. Ensemble learning does not always improve or optimize accuracy. Ensemble learning can improve accuracy by choosing the right base classifier based on each ensemble learning algorithm and characteristics. On the electronic nose dataset for pork adulteration in beef, hard voting with KNN as its base classifier can improve the accuracy from 96.43% to 98.33%. For future works, we will analyze the difference in gas concentration produced by the seven combinations of pork and beef mixtures from the dataset with regression methods.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, Malikhah; methodology, Malikhah; software, Malikhah; validation, Malikhah; writing—original draft preparation, Malikhah; writing—editing, Malikhah; assisted in methodology, Shoffi Izza Sabilla; writing—review, Shoffi Izza Sabilla; supervision, Riyanarto Sarno; proposes problem ideas, Riyanarto Sarno; writing—review, Riyanarto Sarno.

Acknowledgments

The authors would like to acknowledge financial support provided by Indonesia Endowment Fund for Education (LPDP) for funding the research.

References

[1] J. Chairunnisa, T. I. Muda, F. K. Fikriah, R. A. Pramunendar, G. F. Shidik, A. Z. Fanani, and A. Soelaiman, "Pork and Beef Features Extractions", In: *Proc. of International Seminar on Application for Technology of Information and Communication*, Semarang, Indonesia, pp. 295-298, 2018.

[2] R. Sarno, S. I. Sabilla, D. R. Wijaya, D. Sunaryono, and C. Fatichah, "The Dataset for Pork Adulteration from Electronic Nose System", *Data Brief*, Vol. 32, p.106139, 2020.

[3] 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. of International Conf. On Science and Technology*, Yogyakarta, Indonesia, Vol. 1755, pp. 020001, 2016.

[4] S. I. Sabilla, R. Sarno, K. Triyana, and K. Hayashi, "Deep Learning in A Sensor Array System based on The Distribution of Volatile Compounds from Meat Cuts using GC–MS Analysis", *Sensing and Bio-Sensing Research*, Vol. 29, p.100371, 2020.

[5] X. F. Gu, J. W. Xu, S. J. Huang, and L. M. Wang, "An Improving Online Accuracy Updated Ensemble Method in Learning from Evolving Data Streams", In: *Proc. of International Conf. On Wavelet Active Media Technology and Information Processing*, Chengdu, China, pp. 430-433, 2014.

[6] A. Rojarath, W. Songpan, and C. Pong-inwong, "Improved Ensemble Learning for

Classification Techniques based on Majority Voting", In: *Proc. of International Conf. On Software Engineering and Service Sciences*, Beijing, China, pp. 107-110, 2016.

[7] Y. Yang, SC. Chen, and "Ensemble Learning from Imbalanced Data Set for Video Event Detection", In: *Proc. of International Conf. On Information Reuse and Integration*, San Fransisco, USA, pp. 82-89, 2015.

[8] Q. Wang, L. Li, W. Ding, D. Zhang, J. Wang, K. Reed, and B. Zhang, "Adulterant Identification in Mutton by Electronic Nose and Gas Chromatography-Mass Spectrometer", *Food Control*, Vol. 98, pp. 431-438, 2019.

[9] D.E. Pavlidis, A. Mallouchos, D. Ercolini, E.Z. Panagou, and G.J.E. Nychas, "Volatilomics Approach for Off-line Discrimination of Minced Beef and Pork Meat and Their Admixture using HS-SPME GC/MS in Tandem with Multivariate Data Analysis", *Meat Science*, Vol. 151, pp. 43-53, 2019.

[10] S. I. Sabilla, R. Sarno, and K. Triyana, "Optimizing Threshold using Pearson Correlation for Selecting Features of Electronic Nose Signals", *International Journal of Intelligent Engineering and Systems*, Vol. 1, No. 6, pp. 81-90, 2019.

[11] J. Chen, J. Gu, R. Zhang, Y. Mao, and S. Tian, "Freshness Evaluation of Three Kinds of Meats based on The Electronic Nose", *Sensors*, Vol. 19, No. 3, p.605, 2019.

[12] Y. He, X. Bai, Q. Xiao, F. Liu, L. Zhou, and C. Zhang, "Detection of Adulteration in Food based on Nondestructive Analysis Techniques: A Review", *Critical Reviews in Food Science and Nutrition*. pp. 1-21, 2020.

[13] T. Leng, F. Li, L. Xiong, Q. Xiong, M. Zhu, and Y. Chen, "Quantitative Detection of Binary and Ternary Adulteration of Minced Beef Meat with Pork and Duck Meat by NIR Combined with Chemometrics", *Food Control*, Vol. 113, p.107203, 2020.

[14] R. Sarno, K. Triyana, S. I. Sabilla, D. R. Wijaya, D. Sunaryono, and C. Fatichah, "Detecting Pork Adulteration in Beef for Halal Authentication Using an Optimized Electronic Nose System", Vol. 8, pp. 221700-221711, 2020.

[15] H. Liu, Q. Li, B. Yan, L. Zhang, and Y. Gu, "Bionic Electronic Nose Based on MOS Sensors Array and Machine Learning Algorithms Used for Wine Properties Detection", *Sensors*, Vol. 19, No. 1, p. 45, 2019.

[16] Q. Li, F. Yang, L. S. Liu, Z. Z. Zheng, X. J. Lin, and Q. H. Wu, "Classification of Diabetes Disease using TCM Electronic Nose Signals and Ensemble Learning", In: *Proc. of International Conf. On Computer Science & Education*, Vancouver, Canada, pp. 507-511, 2014.

[17] M. A. Bagheri, and G. A. Montazer, "Ensemble Classifier Strategy based on Transient Feature Fusion in Electronic Nose", In: *AIP Conference Proceeding*, New York City, USA, Vol. 1362, pp. 241-242, 2011.

[18] X. Sun, L. Liu, Z. Wang, J. Miao, Y. Wang, Z. Luo, and G. Li, "An Optimized Multi-classifiers Ensemble Learning for Identification of Ginsengs based on Electronic Nose", *Sensors and Actuators A: Physical*, Vol. 266, pp. 135-144, 2017.

[19] M. A. Hasan, R. Sarno, and S. I. Sabilla, "Optimizing Machine Learning Parameters for Classifying the Sweetness of Pineapple Aroma Using Electronic Nose", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 5, pp. 122-132, 2020.

[20] S. Wakhid, R. Sarno, S. I. Sabilla, and D. B. Maghfira, "Detection and Classification of Indonesian Civet and Non-Civet Coffee based on Statistical Analysis Comparison using E-Nose", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 4, pp. 56-65, 2020.

[21] S. I. Sabilla, R. Sarno, and K. Triyana, "Optimizing Threshold using Pearson Correlation for Selecting Features of Electronic Nose Signals", *International Journal of Intelligent Engineering and Systems*, Vol. 12, No. 6, pp. 81-90, 2019.

[22] K.S. Anant, and F.U. Dowla, "Wavelet Transform Methods for Phase Identification in Three-Component Seismograms", *Bulletin of Seismological Society of America*, Vol. 87, No. 6, pp. 1598–1612, 1997.

[23] O. Rioul and M. Vetterli, "Wavelets and Signal Processing", *Wavelets and signal processing*, Vol.8, No.4, pp.14-38, 1991.

[24] R. X. Gao, and R. Yan, *Wavelets: Theory and Applications for Manufacturing*, Springer Science and Business Media, New York, N.Y. 2011.

[25] H. Ankarali, A. C. Yazici, and S. Ankarali, "A Bootstrap Confidence Interval for Skewness and Kurtosis and Properties of T-Test in Small Samples from Normal Distribution", *Medical Journal of Trakya University/Trakya Universitesi Tip Fakultesi Dergisi*, Vol. 26, No. 4, pp. 297-305, 2009.

[26] K. J. Johnson, and R. E. Synovec, "Pattern Recognition of Jet Fuels: Comprehensive GC×GC with ANOVA-based Feature Selection

and Principal Component Analysis", *Chemometrics and Intelligent Laboratory Systems*, Vol. 60, No. 1, pp. 225-237, 2002.
[27] M. H. Dunham, *Data mining introductory and advanced topics*, Pearson Education, 2003.
[28] J. Lever, M. Krzywinski, and N. Altman, "Logistic Regression", *Nat Mehods*, Vol. 13, pp. 541-542, 2016.
[29] H. Bhavsar, and M. H. Panchal, "A Review on Support Vector Machine for Data Classification", *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, Vol. 1, No. 10, pp. 185-189
[30] D. R. Wijaya, R. Sarno, and A. F. Daiva, "Electronic Nose for Classifying Beef and Pork using Naïve Bayes", In: *Proc. of 2017 International Seminar on Sensors, Instrumentation, Measurement and Metrology (ISSIMM)*, Surabaya, Indonesia, pp. 104-108.
[31] P. Sharma, and M. Kaur, "Classification in Pattern Recognition: A Review", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 3, No. 4, pp. 298-306, 2013.
[32] Z. H. Zhou, *Ensemble Methods: Foundation and Algorithm*, CRC Press, 2012.