Feature Selection Based on Minimum Overlap Probability (MOP) in Identifying Beef and Pork

Khoerul Anwar

Departement of Computer Science
and Electronics
Universitas Gadjah Mada
Yogyakarta, Indonesia

Agus Harjoko

Departement of Computer Science
and Electronics
Universitas Gadjah Mada
Yogyakarta, Indonesia

Suharto Suharto

Departement of Computer Science
and Electronics
Universitas Gadjah Mada
Yogyakarta, Indonesia

Abstract—**Feature selection is one of the most important techniques in image processing for classifying. In classifying beef and pork based on texture feature, feature overlaps are difficult issues. This paper proposed feature selection method by Minimum Overlap Probability (MOP) to get the best feature. The method was tested on two datasets of features of digital images of beef and pork which had similar textures and overlapping features. The selected features were used for data training and testing by Backpropagation Neural Network (BPNN). Data training process used single features and several selected feature combinations. The test result showed that BPNN managed to detect beef or pork images with 97.75% performance. From performance a conclusion was drawn that MOP method could be used to select the best features in feature selection for classifying/identifying two digital image objects with similar textures.**

Keywords—*overlap; feature selection; best feature; minimum overlap probability (MOP); identifying*

I. INTRODUCTION

Image identification based on texture features from many images with varying feature types, $X = (x_i, i=1…M)$ is a difficult task. Multi-features of multi-objects for task of certain applications have three probabilities, i.e. relevant, irrelevant, and redundant features [1]. Similarly, according to [2-3], there are three main matters related to feature, i.e., 1) strong relevant, 2) weakly relevant and 3) irrelevant. Based on the argument, not having good knowledge on texture feature will be a problem in determining the best feature which can be used as a key for classification or identification. Therefore, feature selection process is important.

Feature selection is very effective in supporting performance in special tasks [3-7]. Several special tasks in image and computer vision processing are classification [8-10], clustering [11], computational neuroscience, imaging genomics [12 and 13], protein property prediction [14], text mining, image annotation, [15 and 16]. Feature has become an important part in the study of image and computer vision processing [17-18]. Feature is unique identity of an image. Unique identity of an image can be used as a key to recognize an image or can also be used to identify an image from another. Key feature is urgent when certain applications involve hundreds of data with tens of characteristics [1, 12 and 18]. In reality, a feature of two different objects for the same feature with different values is a problem on its own [18].

Feature selection is one of the main tasks in classification. In a large feature collection, it's possible that some or all of them are irrelevant or redundant features. Feature can be collected from extraction on three parts, i.e. texture, shape, and color of object. In this study, feature discussion is focused on texture feature. Feature separation must be performed with the correct extraction because the extraction feature would be used to train classificator. Selecting key feature is a part of a process to improve accuracy in classificator performance [19]. One of the feature selection techniques is selecting features with minimum redundancy criteria in the classification process [7]. Feature selection technique by minimum redundancy is also used by [4] for classification. The main function in selection in choosing minimum redundancy features from two objects which can be used to make distinct classes [2]. However, feature selection of both is used in multi-label features in single object. Unlike previous researchers [2],[4],[8] in this paper the writer performed multi-label feature selection for multi-object on identification of digital images of beef and pork. The selection criteria are features relevant with accuracy of classification. The writer suggested that a feature is relevant when it has minimum overlap. The challenge was choosing one or a few of the feature overlaps as candidates for the best feature. The basic assumption was the smaller the range of overlapping value of a feature, the better it was to select the feature to win the selection (key feature). The basic problem for feature overlap is the feature group isn't a comprehensive representation [20] of the value of a feature as a target because it still contains the value of other features.

In this paper, the writer offered a feature selection to obtain the best features from two groups of features from two digital images with similar textures by **MOP** method. The identity of each feature was determined by min-max values then calculating the overlapping value of each feature. Then overlap probability of each feature was calculated. The next step was selecting key (best) features by applying threshold on probability values. Features with probability less than the threshold won the selection.

This paper was arranged as follows: Section 2, describes relevant scientific works. Section 3, described methodology. Explanation on the MOP the author offered is reviewed in Section 4, Test and Result are described in Section 5 and Discussion in Section 6. Lastly is Conclusion.

II. RELATED WORK

Kamyab and Eftekhari [21] in their paper discuss a special study on the usage of Multimodal Optimization (MO) method for feature selection. To do this, Evolutionary Algorithms (EAs) modification from several famous methods based on and a proposed *niching* method called GA_SN_CM are used for feature selection task and is compared with several famous EA-based methods for feature selection to study the strength of MO method on improving the result of feature selection.

Sotoca and Filiberto [22] in their paper on feature selection, or variable selection, select the most relevant features (attributes) of a group of variable data. In this framework, relevant term refers to the effect of given features or feature set to obtain possibility of minimum error in classification or recognition of classification problems.

Al-Ani, Alsukker and Rami [23] proprose differential evolution algorithm for wrapper feature selection which uses the simplest yet effective way to narrow the search without removing any feature. A number of dataset with different sizes are used to evaluate the performance of the proposed method, which can give good indication on exploration and exploitation of the ability.

Kabir, Islam and Kazuyuki [24] in their paper proposes a new algorithm called constructive approach for Feature Selection (CAFS) based on wrapper approach concept consecutive search strategy. As a learning model, CAFS employs three layers of feed-forward Neural Network (NN). The proposed technique combines feature selection (FS) with NN architecture determination. It uses constructive approach which involves correlation information in selecting features and determining network architecture.

Lutu and Engelbrecht [25] in their paper discuss algorithm for feature selection in data mining prediction for classification problem by trying to categorize them to select relevant and not excessive features for classification task. A relevant feature is defined as one which correlates with target functions. As excessive feature is defined as one which correlates with other features. In this writing, they propose a new algorithm by combining the usage of certain threshold values and decision rules to select feature subset

Hanchuan, Fuhui and Chris [26] combine max-dependency, max-relevance and min-redundancy for feature selection. In reciprocal information, the purpose of feature selection is to discover a set of S features with m feature (x), which together have the biggest dependency in target class c. This scheme is called max-dependency. But it's difficult to do. The alternative is max-relevance which is searching for features which fulfill max-dependency value by average value of all reciprocal information values between features of individual x and class.

III. FRAME WORK MOP

Feature selection was performed to get the best features from feature set of beef and pork for classification task. Every feature set has 20 features where every feature consists of 200 data. The problem was the values of extraction features between beef and pork from the same feature name didn't

produce independent features, but feature overlaps instead. Fig 1 shows the framework for feature selection by MOP method. The focus on this stage is making a model to calculate area overlap of every feature between features of digital images of beef (*Fs*) and pork (*Fb*. The early stage is determining the area of each feature of *Fs* and *Fb*. It's continued by calculating the area overlap of both. Then, probability value of every feature was calculated. Lastly, overlap probability value was selected. The constructed model architecture is shown in Fig 1.
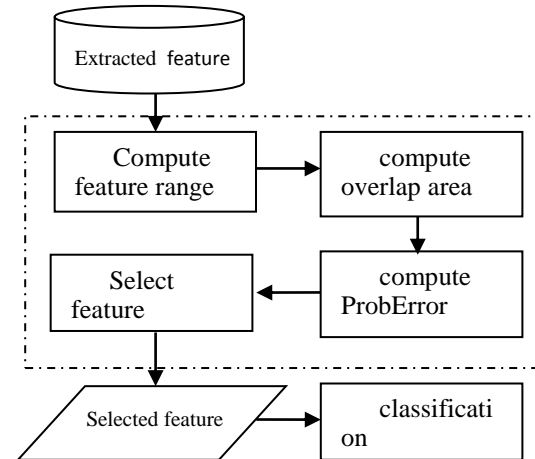

Fig. 1. Framework of feature selection by MOP

To determine whether selected features are the best features as expected, selected features were tested on artificial neural network (ANN).

A. Extraction featur

Extraction is a pre-processing stage which is a basic stage to get maximum data before processing. Extraction was performed on each image to determine texture characteristics. The feature which became the object of the writer's research was the feature of the texture of digital images of beef and pork. The main thing to get was strong features which could be used to differentiate the texture of both. As usual, to get features in pre-process, the study conducted extraction process of both by several types of features which have been used by previous researchers. Some of those features were used to look for unique features from the extracted images. The extraction model in this study was gray level co-occurrence matrix (GLCM) method. GLCM is a tabulation of how often different combinations of gray level co-occurrence matrix are found in image section or images [27]. Calculation of texture feature used GLCM to get sizes of variations in intensity (i.e, image texture) in pixels which were focused on. Co-occurrence matrix was calculated by two parameters, which is relative distance between d of pixel pair measured in total pixel and their relative θ orientation. These two parameters were expected to find special characteristics of two digital images of beef and pork. Unique features expected to be found maximally from the digital images were: autocorrelation, contrast, correlation, cluster prominence, cluster shade, dissimilarity, energy, entropy, homogeneity, maximum probability, sum of square variance, sum average, sum variance, sum entropy, difference variance, difference entropy, information of correlation, Inverse difference normalized,

inverse difference moment normalized. The extraction result was numeric. The numbers were the data or sources of data processing. The extraction result showed the group of feature values for digital images of beef $S = (x_1, x_2, x_3 \dots x_n)$ and for digital images of pork, $B = (x_1, x_2, x_3 \dots x_n)$, with x being feature names. x feature has a group of value from extraction of a number of extracted images. The names and formulas of extraction features used for classification/identification of beef and pork in this research was cited from [28] [29], as shown in Table I

TABLE I. EXTRACTION FEATURE

No	Features name	Formulation
1	Autocorrelation	$\sum_{i=0}^{n} \sum_{j=0}^{n} (i.j) P_{r,\emptyset}^t(i,j)$
2	Contrast	$\sum_{n=0}^{n_c-1} n^2 \left\{ \sum_{i=1}^{n_c} \sum_{j=1}^{n_j} p(i,j), |i-j| = n \right\}$
3	Correlation	$\dfrac{\sum_{i=0}^{n} \sum_{j=0}^{n} (i - \mu_{x,r,\theta}^t)(j - \mu_{y,r,\theta}^t) P_{r,\theta}^t(i,j)}{\sigma_{x,r,\theta}^t \sigma_{y,r,\theta}^t}$
4	Cluster Prominence	$\sum_{i=0}^{n} \sum_{j=0}^{n} ((i - \mu_{x,r,\theta}^t) + (j - \mu_{y,r,\theta}^t))^4 P_{r,\theta}^t(i,j)$
5	Cluster Shade (skewness)	$\sum_{i=0}^{n} \sum_{j=0}^{n} ((i - \mu_{x,r,\theta}^t) + (j - \mu_{y,r,\theta}^t))^3 P_{r,\emptyset}^t(i-j)$
6	Dissimilarity	$\sum_{i=0}^{n} \sum_{j=0}^{n} (i-j) P_{r,\emptyset}^t(i,j)$
7	Energy	$\sum_{i=0}^{n} \sum_{j=0}^{n} P_{r,\theta}^t(i,j)^2$
8	Entropy	$\sum_{i=0}^{n} \sum_{j=0}^{n} P_{r,\theta}^t(i,j)^2 log_2(P_{r,\theta}^t(i,j))$
9	Homogeneity	$\sum_{i=0}^{n} \sum_{j=0}^{n} \dfrac{1}{1+(i-j)^2} P_{r,\emptyset}^t(i,j)$
10	Maximum probability	$\dfrac{argmax}{i,j} P_{r,\theta}^t(i,j)$
11	Sum of sqaures: Variance	$\sum_{i} \sum_{j} (i - \mu)^2, p(i,j)$
12	Sum average	$\sum_{i=2}^{2N_c} i p_{x+y}(i)$
13	Sum variance	$\sum_{i=2}^{2N_c} (i - S_{ent})^2 P_{x+y}(i)$
14	Sum entropy	$-\sum_{i=c}^{2N_c} p_{x+y}(i) log (P_{x+y}(i)) = S_{ent}$
15	Difference variance	$\sum_{i=0}^{N_c-1} i^2 p_{x-y}(i)$
16	Difference entropy	$-\sum_{i=c}^{N_c-1} p_{x-y}(i) log (P_{x-y}(i))$
17	Information measure of correlation1	$\dfrac{HXY - HXY1}{max\{HX, HY\}}$ HX, HY adalah entropy pada px dan py , sehingga $HXY = -\sum_{i=0} \sum_{j=0} P(i,j) \ log(P(i,j))$ $HXY1 = -\sum_{i=0} \sum_{j=0} P(i,j) \ log(P_x(i)P_y(j))$

No	Features name	Formulation
		$HXY2 = -\sum_{i=0} \sum_{j=0} P_x(i) P_y(j) log(P_x(i)P_y(j))$
18	Informaiton measure of correlation2	$(1 - exp| - 2(HXY2 - HXY)|)^{1/2}$
19	Invers Difference Normalized	$\sum \dfrac{C(i,j)}{1+|i-j|}$ Dimana $C(i,j) = \dfrac{p(i,j)}{\sum_{ij=1}^{\theta} p(i,j)}$
20	Invers Difference Moment Normalized	$\sum \dfrac{C(i,j)}{1+(i-j)^2}$

B. Feature Range

Based on the extraction, the value of extraction features for 20 type of the features showed there was no particular feature which has independent range or categorized as strong relevance. Instead there were overlaps. So, in this study, the value of feature overlap received special attention or led to further study. This problem required certain formulation which can be used to determine the values of features in overlap area [16]. This method aims to get features with minimum overlap range probability and select features by certain thresholds to get the best features. Feature range is a range formed by minimum and maximum values. The formula was as follows:

1) Max Value: Maximum value is the highest value of a feature of a data set.

$$Max_x = max (x_1:x_m) \qquad (1)$$

2) Min Value: Minimum value is the lowest value of a feature.

$$Min_x = min (x_1:x_m) \qquad (2)$$

3) Feature Range: Area marked by minimum and maximum value limits of a feature

$$Fitur_x = (min_x:max_x) \qquad (3)$$

$x_1;x_m$ is a group of value of the 1st x feature to *m* amount of data

Ranges of beef features (*Fs*) and pork features (*Fb*) were areas formed by Min_x and Max_x (3) of extracted data. These min-max values were respectively used as the lower limits and upper limits of feature areas. So the area of every *Fs* and *Fb* could be determined. To determine overlapped area visually between *Fs* amd *Fb* when interacting, each feature area was visualized in two dimensions (2D). The feature areas could be formed by giving range values of features to *x* and *y* axis. Therefore, value range in *x* axis was (*min,max*), the same value range applied to *y* axis. When similar features (*X*) from beef and pork image data were described in the field, the features could be analyzed.

Fb Range, Ordinate point (x_1,y_1) was the lower left corner point or equaled to feature value $x_1(min.min)$, (x_2,y_1) was the lower right corner point which equaled to the feature value $x_1(max.min)$, (x_1,y_2) was the upper left corner point equaled to feature value $x_1(min,max)$ and (x_2,y_2) was the upper right

corner point equaled to feature value $x_1(max,max)$. *Fb* area could be determined based on the ordinates, so *Fb* was

$$Fb = ((x_1,y_1),(x_2,y_1) ; (x_1,y_2),(x_2,y_2)) \quad (4)$$

Fs Range, Ordinate point (x_3,y_3) was the lower left corner point or equaled to feature value $x_1(min,min)$, (x_4,y_3)) was the lower right corner point which equaled to the feature value $x_1(min,max)$, (x_3,y_4) was the upper left corner point equaled to feature value $x_1(max,min)$ and (x_4,y_4) was the upper right corner point equaled to feature value $x_1(max,max)$. *Fs* area could be determined based on the ordinates, so *Fs* was

$$Fs = ((x_3,y_3),(x_4,y_3);(x_3,y_4),(x_4,y_4)) \quad (5)$$

C. Overlap

Overlap between Fs happen Fb happened when they surpassed the value ranges of two or more features. In this study, feature had value range (3). Based on (3) $Fs(X_1)$ and $Fb(X_1)$ features overlapped when the maximum values of $Fs(X_1)$ were bigger than the minimum values of $Fb(X_1)$ and the minimum values of $Fs(X_1)$ were less than the minimum values of $Fb(X_1)$. In theoretical discussion, overlapping set is called intersection. The formulation to get intersection value was

$$Fs \cap Fb = \{x \mid x \in Fs \} \cap (x \mid x \in Fb\} \quad (6)$$

Equation (5) was effective to determine intersection element. However, this study was aimed to determine range, so this study modified (6) to find intersection value. The term intersection in this study was called overlap.

Feature overlap area *(Fo)*: *Fo* was overlapping area of *Fb* and *Fs* areas. This area was determined by the positions of ordinates *Fb* (4) and *Fs* (5). *Fo* area could be determined by ordinates *Fs* and *Fb*, so *Fo* area was

$$Fo= ((x_3,y_3),(x_2,y_3);(x_3,y_2),(x_2,y_2)) \quad (7)$$

Ordinate points in (7) were overlapped ordinate points which were, respectively, (x_3,y_3) which equaled to feature value *Fs* fpr *(min,min)*, (x_2,y_3) was feature values *Fs* and *Fb* *(Fs min, Fb max)*, (x_3,y_2) equaled to feature values *Fs* and *Fb* *(Fb max, Fs min)*, and (x_2,y_2) equaled to feature value *Fs* for *(max,max)*. Fig 2 shows overlap area between a feature of *Fs* and Fs.
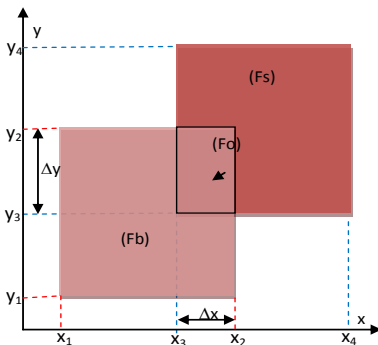

Fig. 2. Overlap area between Fb and Fs

Fig 2 is interaction of *Fs* and *Fb* which shows an overlap between *Fs* and *Fb*.

D. MOP

Probability of an incidence is a number which shows the possibility of an event. In this study, there was possibility of similar value between *Fs* and *Fb* in overlap area. The problem was how big the overlap was between them. Using set theory, the number of Fs members could be written as nFs, and the number of Fb set members could be written as nFb. Based on this, *probability of overlap area* (ProbArea) was defined as

$$ProbArea = \frac{|nFs \cap nFb|}{|nFs \cup nFb|} = \frac{|nFs \cap nFb|}{|nFs+nFb|-|nFs \cap nFb|} \quad (8)$$

In this study, the author modified (8) for the number of members of sets with feature area size. The formulation of each size is defined below:

Size of feature areas of beef (Ls) and pork (Lb): The general formula of area size is length multiplied with width. Length in this case the length of *Fs* was the range of *Fs* along X axis, i.e. delta (Δx_s) the distance between maximum point (x_4) and minimum point (x_3). While width of *Fs* was the range of *Fs* along Y axis, i.e. delta (Δy_s) the distance between maximum point (y_4) and minimum point (y_3). While the length of *Fb* was the range of *Fb* along X axis, i.e. delta (Δx_b) the distance between maximum point (x_2) and minimum point (x_1). While the width of *Fb* was the range of *Fb* along Y axis, i.e delta $(\Delta y_b$ the distance between maximum point (y_2) and minimum point (y_1). So, *Ls* was defined as:

$\Delta x_s = |x_4 - x_3|$

$\Delta y_s = |y_4 - y_3|$

$$\Delta x_s = \Delta y_s \quad (9)$$

where

$$Ls = |x_4 - x_3| * |y_4 - y_3| \quad (10)$$

Based on (9) and (10), so

$$Ls = \Delta x_s^2 = \Delta y_s^2 \quad (11)$$

and *Lb* could be defined as:

$\Delta x_b = |x_2 - x_1|$

$\Delta y_b = |y_2 - y_1|$

$$\Delta x_b = \Delta y_b \quad (12)$$

where

$$Lb = |x_2 - x_1| * | y_2 - y_1| \quad (13)$$

Base on (12) and (13), so

$$Lb = \Delta x_b^2 = \Delta y_b^2 \quad (14)$$

1) Size of overlap area (Lo): To determine the length and width of *Lo* area, Fig 2 shows the length of *Lo* is Δx and the width Δy. Each could be calculated by the following equations:

$$\Delta x = |x_2 - x_3| \quad (15)$$
$$\Delta y = |y_2 - y_3| \quad (16)$$

Equations (15) and (16), for $\Delta x = |x_2 - x_3|$, point x_2 was maximum value for *Fb* and point x_3 was minimum value of *Fs*. $\Delta y = |y_2 - y_3|$, point y_2 x_2 was maximum value for *Fb* and point y_3 was minimum value of *Fs*. So based on (15) and (16) the size of overlap area (*Lo*) was

$$Lo = \Delta x * \Delta y \quad (17)$$

Based on (11) and (14), *Ls* and *Lb* were areas with the same length on both sides. So *Lo* in (17), could be written as

$$Lo = \Delta x^2 \qquad (18)$$

Using *Lo* (18) for feature values in this area was indicated to cause problem for the process of identifying images of beef or pork. The problem happened in the area maybe due to duplication of values of features of beef and pork. Thus, the bigger the value of *Lo*, the bigger the amount of duplication of members of features of beef and pork, and vice versa. It should be noted that this area was formed by the range of feature values, so the overlap area of every feature wasn't absolute at certain amounts because the range was influenced by the area stability of each feature. However, by using area range based on min-max of the features, the system was still able to get overlap area. The possibility of overlap or overlap probability was the main focus of this study. To determine overlap probability of every feature between *Fs* and *Fb, Lo* could be compared with the size of all features (Fs + Fb). In this study, computation of overlap probability by the author was called error probability (*ProbError*). The formula was

$$ProbError = 2 * Lo / (Ls + Lb) \qquad (19)$$

Equation (19) meant that the smaller the value of *Lo,* the smaller the value of *ProbError*. Conversely, the bigger the value of *Lo,* the bigger the value of *ProbError*.

The author used this ProbError value as data to select the best features. Selection was performed by giving threshold value < 10%. The author named this method *Minimum Overlap Probability (MOP)* method.

E. MOP Algorithm:

1) Algorithm of feature selection by *Minimum Overlap Probability (MOP)* method

a) Calculating minmax values of extraction features of digital images of beef and pork

b) Determining feature area (F_x) of digital images of beef and pork.

- If F_x had no overlap (independent) it's a selected feature

- If F_x was a subset or superset of each other, F_x wasn't a selected feature (rejected)

- If the calculation delta *Fx* of a group of feature database wasn't in *a* or *b* process, the process was continued (process 3).

c) Calculating ProbError value

d) Determining threshold (as filter of selection of selected features)

e) Finding features with ProbError less than treshold

f) Selected features

F. MOP Flowchart

Algorithm of feature selection by MOP method is illustrated in the flowchart in Fig 3.
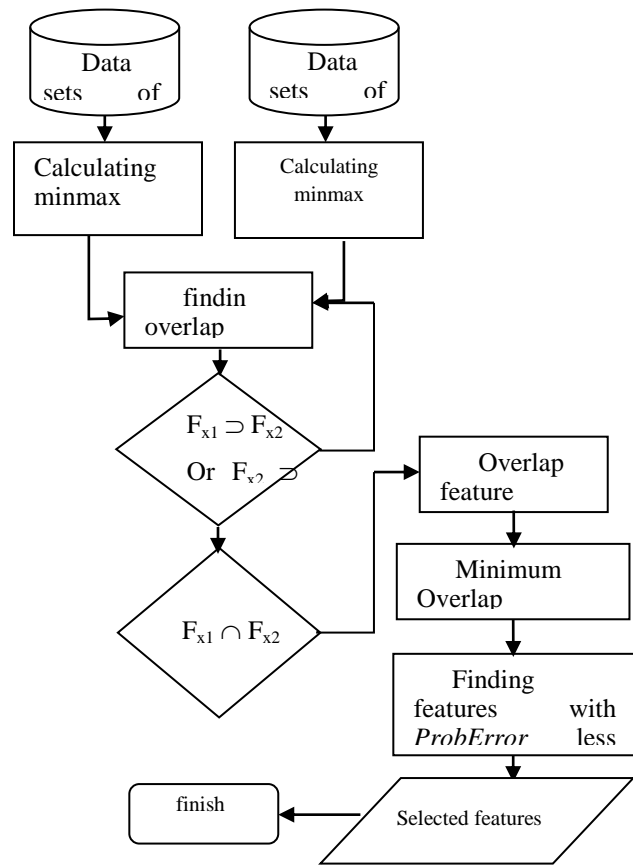

Fig. 3. Flowchart of Feature selection by MOP

G. Testing the selected features

Testing the features selected conducted on artificial neural networks. This testing is done to determine the effect on the accuracy of results. Type of neural network used is a multi-layer back propagation neural network.. The network architecture used here was I-H-O i.e. input layer, hidden layer and lastly layer output. To determine the correlations of selected features and accuracy of network classification, input layer was set up for several nodes. Meanwhile, output layer was set up was two nodes. In training stage, the target classes were label 00 for pork, 11 for beef. To support the performance of the network, the selected learning method was levenberg marguad. It's because this method has the best accuracy compared with other learning methods.

IV. EXPERIMENT AND RESULT

A. Experiment

Image data was acquired by mobile digital camera at five mega pixel (5MP). Total data was 400 images with 200 images each for beef and pork. The dimension of digital images was 255 x 255 and in JPEG format. Data was pre-processed by converting RGB to gray, filtering images by gabor. Total and names of extraction features are written in Table I. Treshold value was 10%. The process in this experiment is as follows:

1) Reading Fb and Fb data sets

2) Determining the range of min and max values of Fs and Fb (3)

3) Calculating proberror value by equation (14).

4) Selecting features. In this selection process, a criterion was used to select the best features. The best features met the following criterion:

$$f(x) = \begin{cases} 1, & Proberror < treshold \\ 0, & Proberror \geq treshold \end{cases} \qquad (15)$$

Note (15) f(x) is selected feature, threshold < 10%. Criterion f(x) =1 means fulfilling requirement or accepted, while f(x)=0 means not fulfilling the criterion or rejected.

5) The final step was testing selected features on neural network.

B. Result

The range values of Fs and Fb, and error probability of every feature from the implementation of MOP method produced the result shown in Table II

TABLE II.   RANGE VALUES AND ERROR PROBABILITY

Feature Name	Fs Range		Fb Range		Prob Error
Maximum probability	0.61	0.76	0.74	0.95	0.06
Contrast	0.12	0.44	0.02	0.15	0.07
Difference variance	0.12	0.44	0.02	0.15	0.07
Sum average	2.51	3.10	2.13	2.58	0.07
Autocorrelation	1.95	3.32	1.35	2.10	0.08
Energy	0.39	0.60	0.56	0.90	0.08
Sum entropy	0.85	1.37	0.25	0.94	0.08
Entropy	0.92	1.67	0.26	1.05	0.09
Dissimilarity	0.11	0.32	0.02	0.14	0.10
Sum of sqaures: Variance	1.92	3.49	1.33	2.14	0.10
Sum variance	4.28	7.07	4.03	4.59	0.10
Roughness	0.05	0.12	0.03	0.06	0.11
Difference entropy	0.34	0.72	0.10	0.41	0.11
Informaiton measure of correlation2	0.63	0.74	0.44	0.67	0.13
Homogeneity	0.85	0.95	0.93	0.99	0.14
Cluster Shade (skewness)	11.43	23.66	10.67	13.29	0.14
Homogeneity	0.86	0.95	0.93	0.99	0.15
Inverse difference normalized (INN)	0.97	0.99	0.98	1.00	0.33
Correlation	0.80	0.90	0.84	0.96	0.38
Correlation	0.80	0.90	0.84	0.96	0.38
Information measure of correlation1	0.32	0.72	0.09	0.42	0.16
Regularity	0.00	0.03	0.01	0.04	0.69
Inverse difference moment normalized	0.99	1.00	1.00	1.00	100.00
Cluster Prominence	135.56	230.61	135.54	170.01	Error

On Table II, feature selection was performed by using the determined filter value (threshold < 10%). The result of the selection was names of selected features shown in Table III.

TABLE III.   SELECTED FEATURES

No	Feature name	Fs Range		Fb Range		Prob Error (%)
1	Maximum probability	0.61	0.76	0.74	0.95	6
2	Contrast	0.12	0.44	0.02	0.15	7
3	Difference variance	0.12	0.44	0.02	0.15	7
4	Sum average	2.51	3.10	2.13	2.58	7
5	Autocorrelation	1.95	3.32	1.35	2.10	8
6	Energy	0.39	0.60	0.56	0.90	8
7	Sum entropy	0.85	1.37	0.25	0.94	8
8	Entropy	0.92	1.67	0.26	1.05	9

Table III shows selection result with error probability values of 6%,7%,8% and 9%. Some features had the same error probability values, e.g. contrast, difference variance, sum average with error probability value of 7%, autocorrelation, energy, sum entropy with error probability value of 8%.

The test on selected features for classification task was performed on neural network. The architecture of neural network was five nodes of input, five nodes of hidden layer and 2 nodes of output (5I-5H-2O). There were two models of the test, first using the best feature as single input and second using combination of some of the best features. The performance of neural network for data training by best feature input (maximum probability) produced 95.50%. A different result was shown by several combinations of the best features which produced100%. The result of data testing by some combinations of features as input of neural network was shown in Table IV.

TABLE IV.   ACCURACY OF FEATURE COMBINATION ON NN

Combinations of features	accuracy (%)
1,2	96.75
1,2,3	96.00
1,2,3,4	95.00
1,2,6,8	97,75
1,2,3,4,5	92.75
1,2,3,4,5,6	95.50
1,2,3,4,5,6,7	97.50
1,2,3,4,5,6,7,8	94.75

Table IV is the result of data testing based on data classification. Combination of features 1,2,6,8 had the highest accuracy of neural network (97.75%), while combination of features 1,2,3,4,5 had the lowest accuracy (92.75%). It showed that classification by combination of selected features produced accuracy of performance of neural network above 92.00%

V.   DISCUSSION

Feature selection by MOP with threshold 0.1 selected maximum probability, contrast, Difference variance, Sum average, Autocorrelation, energy, Sum entropy, and entropy as the best features from 20 feature candidates. It meant that

these features in digital images of beef and pork had smaller overlap values than other features. Combinations of selected features were used to train network and then testing was performed using new data, showing the best features could support network performance. The lowest network accuracy was in feature combination 1,2,3,4 and 5 with 92.75% accuracy or error level of 7,25%. The best accuracy was in feature combination 1,2,6 and 8 which had network performance with 97.75% accuracy or error level of 2.25%. It showed that feature combinations influenced accuracy of classification. Based on the result of the test, the selected features were correct and could be used as unique characteristics to identify beef or pork by digital image.

VI. CONCLUSION

Overlap probability can be used to select the best features of some of the features that have value overlap one another. MOP method could be used as one of the solutions for selecting the best or strongest features of two objects with feature overlap.

The selected feature is a maximum probability, contras, energy and entropy is the best feature based on the results of testing with artificial neural networks. It is derived from the performance of the neural network with an accuracy rate of 97.75%. In other perngertian error rate of of 2.25%.

Future work will be the development Minimum Overlap Probability method to determine the correlation between the selected feature.

REFERENCES

[1] B. Xue, M. Zhang, and W. N. Browne, "Particle Swarm Optimization for Feature Selection in Classification: A Multi-Objective Approach." *IEEE Transactions On Cybernetics*, pp 1656-1671, 2013.

[2] J.R. Vergara, and P. A. Estevez, "A review of feature selection methods based on mutual information." *Neural Comput & Applic* xxiv, pp 175–186, 2014.

[3] L.Huawen, W. Xindong, and Z. Shichao, "A new supervised feature selection method for pattern classification." *Computational Intelligence* (Blackwell Publishing Inc) XXX, No. 2, pp 342-361, 2014.

[4] L.Yaojin, H. Qinghua, L.Jinghua, and D. Jie, "Multi-label feature selection based on max-dependency and min-redundancy." *Neurocomputing* ,Elsevier, pp 92–103, 2015.

[5] H. Zuhair, A.Selmat, and M.Salleh, "The Effect of Feature Selection on Phish Website Detection." *(IJACSA) International Journal of Advance Computer Science an Applications* VI, No. 10, pp 221-232, 2015.

[6] S. Elseuofi, and W. Awad, "Toward Accurate Feature Selection Based on BSS-GRF." *(IJACSA) Intenational Journal of Advance Computer Science an Applications* V, No. 8, pp 55-59, 2015.

[7] P. Hanchuan, L. Fuhui, and D. Chris, "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance,and Min-Redundancy." *IEEE Transactions On Pattern Analysis And Machine Intelligence,* IEEE, Vol 27, No. 8 pp 1226-1238, August 2005.

[8] C. Bolun, C. Ling, and C. Yixin, "Efficient ant colony optimization for image feature selection." *Signal Processing*, pp 1566–1576, 2013.

[9] L. Xiaoming, and T. Jinshan, "Mass Classification in Mammograms Using Selected Geometry and Texture Features, and a New SVM-Based Feature Selection Method." *Systems Journal, IEEE* ,Vol. 8, No. 3,pp 910-920, *September 2014.*

[10] Y. V. Goncharov, I. B. Muchnik, and L. V. Shvartser, "Feature Selection Algorithm in Classification Learning Using Support Vector Machines." *Computational Mathematics and Mathematical Physics*, pp 1243–1260,2008.

[11] A. Emre, and K. Adem, "Simultaneous feature selection and ant colony clustering." *Procedia Computer Science,* Elssevier, pp 1432–1438, 2011.

[12] W.Hua, et al. "Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the ADNI cohort." *Bioinformatics,*Vol 28 , No. 2, pp 229–237,2012.

[13] W. Awada, T.M. Khoshgoftaa, D.Dittman, R. Wald, and A. Napolitano, "A Review of the Stability of Feature Selection Techniques for Bioinformatics Data." IEEE, pp 356-363, 2012.

[14] W. De, Y. Liang, F. Zhengqi, and X. Jingbo, "Prediction of Thermophilic Protein with Pseudo Amino Acid Composition: An Approach from Combined Feature Selection and Reduction." *Protein And Peptide Letters*, 684-689, March 2011.

[15] D. D Lewis, "Feature selection and feature extraction for text." *Procceding. Workshop Speech Natural Language.* Stroudsburg, pp 212–217, 1992.

[16] D. Wang, F.Nie, and H. Huang, "Feature Selection via Global Redundancy Minimization." *IEEE Transactions On Knowledge And Data Engineering,*Vol 27, No. 10, pp 2743-2755, October 2015.

[17] S. An, and X. Fan, "Study on Method of Feature Selection in Speech Conent Classsification." *(IJACSA) International Journal of Advanced Computer Science and Applications,* Vol 54, No. 4 pp 71-75, 2014.

[18] S.N Devi, and S.P Rajagopalan. "A study on Feature Selection Techniques in Bio-Informatics." *(IJACSA) International Journal of Advance Computer Science dan Applications, Vol* 2, No. 1, pp 138-144, 2011.

[19] J. Liu, and G. Wang, "A Hybrid Feature Selection Method for Data Sets of thousands of Variables." *IEEE*, pp 288-291, 2010.

[20] N. Valliammal, and S.N. Geethalakshmi, "An Optimal Feature Subset Selection for Leaf Analysis." *International Journal of Computer, Electrical, Automation, Control and Information Engineering,* Vol 6, No. 2, pp 191-196, 2012.

[21] C. Ding, and H. Peng, "Minimun Redundancy Feature Selection From microarray Gene Expression Data." *Journal of Bioinformatics and Computational Biology*, pp 185-205, 2005.

[22] Kamyab, Shima, and Mahdi Eftekhari. "Feature selection using multimodal optimization techniques." *Neurocomputing,* pp 586–597, 2016.

[23] J.M. Sotoca, and P. Filiberto, "Supervised feature selection by clustering using conditional mutual information-based distances." *Pattern Recognition,* Elsevier, Vol 43, pp 2068–2081, 2010.

[24] A. Al-Ani, A. Alsukker, and N. K. Rami, "Feature subset selection using differential evolution and a wheel based search strategy." *Swarm and Evolutionary Computation,* Elsevier, pp 15–26, (2013).

[25] M.M Kabir, M.MIslam, and M. Kazuyuki, "A new wrapper feature selection approach using neural network." *Neurocomputing,* Elsevier, vol 73, pp 3273–3283, 2010.

[26] P.E. Lutu, and A.P. Engelbrecht, "A decision rule-based method for feature selection in predictive data mining." *Expert Systems with Applications,* Elsevier, Vol 47, pp 602–609, 2010.

[27] H. Shahera, and S.Seiichi, "Features for Texture Analysis." *SICE Annual Conference.* Akita, Japan: SICE, pp 1939-1744, 2013.

[28] L.K Soh, and C. Tsatsoulis. "Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices." *Transactions on Geoscience and Remote Sensing, IEEE, Vol* 37 pp 780-795, March 1999.

[29] D.A. Clausi, and Y. Zhao, "Grey level co-occurence integrated algorithm (GLCIA): a superior computational method to rapidly determine co-occurence probability texture features." *Computer & Geosciences*, pp 837-850, 2003.