Feature Selection on Facial Expression Recognition System using Low Variance Filter

Eldio Ruben Marshall Langkun[1], Febryanti Sthevanie, S.T., M.T.[2], Imelda Atastina, SSi., M. T.[3]

[1,2,3]Informatics Faculty, Telkom University, Bandung
[1]eldiolangkun@students.telkomuniversity.ac.id, [2] sthevanie @telkomuniversity.ac.id,
[3]imelda@telkomuniversity.ac.id

Abstract
Facial expression recognition can differentiate human's current state of emotion, representing what a person is feeling. This can be implemented in many areas such as mental state analysis, human-computer development, etc. Current research already produced good accuracies but not paying more attention to the number of data or performance. This research aimed to produce facial emotion recognition system that focused in reducing the number of features based on facial expression images. We differentiate seven human emotion using Biorthogonal Wavelet Transform(BWT) to extract the features from the image, and utilize Low Variance Filter to reduce high features, and based on the extracted data we build classification model using Support Vector Machine(SVM), also stratified cross validation is used as the validation model. The proposed model successfully produced in average accuracy of 92% even after 95% data reduction is applied to the model. In short, this proposed method is efficient.

Keywords: Facial Expression Recognition, Biorthogonal Wavelet Transform, Low Variance Filter

1. Introduction

1.1. Background

Feature selection as a method of dimension reduction is crucial in building a classification model. It can extract the important information from the data, focusing the model to that information and extract other unessential data, preventing problems like overfitting. Especially in expression recognition that is key to human-computer interaction [1] [2]. A feature selection can extract important information in expression recognition like the facial muscle condition that differentiate between expressions, deleting other unimportant information like the background or human hair.

There are many research facial expression recognition based on facial images that includes dimension reduction. A facial emotion recognition experiment using Biorthogonal Wavelet Entropy(BWE) and Fuzzy Support Vector Machine (FSVM) [3] resulting an overall accuracy up to 96.77%. In 2010, a FER research using Local Binary Pattern(LDP) and Principal Component Analysis(PCA) [4] producing 96% accuracy to differentiate 6 human expression Those researches produce a magnificent accuracy, but even they include a dimension reduction method, there are no further analysis on the effect to data reduction for the dataset.

This research is focused on the affection of feature selection to reduce the dataset for a facial expression recognition system. Considering the most important features in expression image is the facial muscle that are recognized by edge detection, we choose Low Variance Filter as the feature selection method. Low variance filter has applied as edge-detection method and has proofed to be effective compared to other method [5] that will serve as a reduction dimension method.

We design the model by first implementing Biorthogonal Wavelet Transform (BWT) because of its ability to recognize fine details and coarse details in multilevel analysis [6], producing great quality of detected edge and combining it with Low Variance Filter. To build the classification model, we use the well-known Support Vector Machine(SVM) with OneVsRest Multiclass approach.

1.2. Problem Statement

Based on the background that has been explained before, therefore we can identify the problems in this research, which is as follows:

1. How the Low Variance Filter will influence the classification model that designed using Biorthogonal Wavelet Transform and Support Vector Machine?

To decrease of this research scope, where must be boundaries to the research which as follows:

1. Model recognize up to seven emotion which is happy, sad, angry, disgust, scared, surprised and neutral
2. Input image is a front look facial female human image without any accessories
3. Support Vector Machine is the only classification method used.

1.3. Goals

The objective of this research is to analyse the affection of Low Variance Filter on the classification model that are build using Biorthogonal Wavelet Transform and Support Vector Machine.

1.4. Paper Structure

The remaining part of this paper is as follows. Section 2 gave the literature review and materials that are used. Section 3 describe about the methodology in the research. Section 4 is the analysis of the results that are produced, and finally section 5 concludes the paper.

2. Literature Review

Table 1. literature review important points

No	Paper Title	Topic	Method	Result
1	Facial Emotion Recognition Based on Biorthogonal Wavelet Entropy, Fuzzy Support Machine, and Stratifiend Cross Validation [3]	Facial Expression Recognition	Biorthogonal Wavelet Entropy and Fuzzy Support Vector Machine	BWE+FSVM proofed to be effective classification model with overall accuracy of 96.77% better than other state-of-the-art methods.
2	Robust Facial Expression Recognition Based on Local Directional Pattern (LDP) [4]	Facial Expression Recognition	Local Directional Pattern for feature extraction, PCA or AdaBoost for reduction dimension and SVM for classification.	LDP proves to be superior than LBP and Gabor filter, with 95% accuracy. Also it proves that the number of feature could affect the accuracy.
3	Variance Filter for Edge Detection and Edge-Based Image Segmentation [5]	Edge-Detection	Variance Filter used for detecting edges in an image	In case of low contrast and significantly blurred edges the introduced method outperforms traditional gradient based approaches to edge detection.

2.1. Biorthogonal Wavelet Transform

Wavelet Transform is a method of feature extraction that uses wavelet or "little wave" that have the ability to represent time and frequency simultaneously [7]. Wavelet transform is chosen because if it's ability to recognize fine and coarse edges in multi-level decomposition. Biorthogonal wavelet type is chosen because of its superiority over orthogonal wavelet [8]. Mathematically, in BWT a signal *x(n)* is implemented a series of low-pass filter *l* and high-pass filter *h* and through a downsample that resulting approximation coefficient C_l and detail coefficient C_h as shown in formula 2.1 and 2.2.

$$C_l(n) = \sum_{m=-\infty}^{+\infty} x(m) \times l(2n-m) \tag{2.1}$$

$$C_h(n) = \sum_{m=-\infty}^{+\infty} x(m) \times h(2n-m) \tag{2.2}$$

In 2D image processing, instead of two coefficients, each level of BWT will produce four coefficient subbands which is called approximation image(C_a), horizontal edges(C_h), vertical edges(C_v), and diagonal edges(C_d) that are visible in figure 1. In Figure 1(a-b) indicates the multilevel capability or the further decomposition of the 2D BWT, revealing more details as the level of decomposition arises.

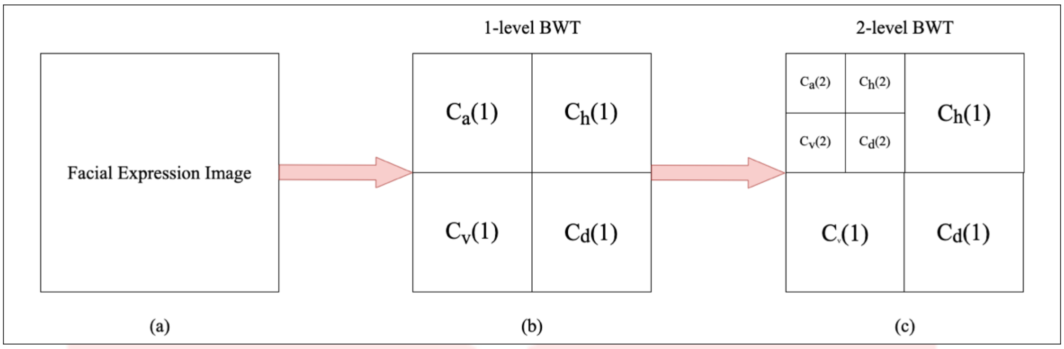


Figure 1. (a) Facial expression image; (b)1-level BWT; (c) 2-level BWT

2.2. Low Variance Filter

The variance filter produces a calculation of local similarity in an image, it focuses in sudden changes in image brightness [9]. Low variance filter has the ability to recognize areas that has no specific changes between its elements based on its variance and threshold. This means that based on the input threshold, the filter is effective to recognize less important elements in the data set like the background of the image and reduce it while at the same time preserving essential information like the image edges, which hypothetically will perform well with feature extraction method that extract edges like biorthogonal wavelet transform.

Suppose we have n by n pixels then we can exclude a set from that data to calculate its variance. The variance of such a set is given in formula 2.3.

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{2.3}$$

Where n is the number of pixels in the set, x_i is the value of i-th pixel, and $\bar{x}$ is the average of pixel values in the given set [5].

2.3. Support Vector Machine

The support vector machine (SVM) purpose is to find the best hyperplane that means having the highest margin between two class of classification [10]. Margin itself can be defined as the distance of hyperplane with point closest to each respective classification class.

At first, SVM is created to classify two types of class only. But, in this case, we are dealing with multiclass classification where we classify seven different type of classes, that is why we use the winner-takes-all (WTA) or

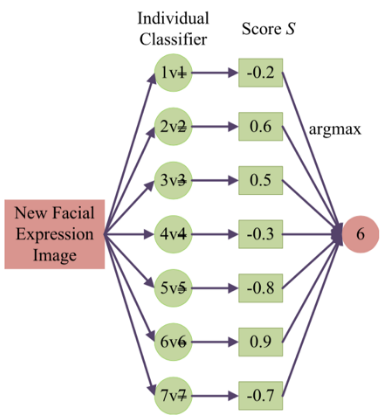


Figure 2. Seven classification class using OVR approach [4]

OneVsRest (OVR) approach as shown in figure 2. Basically, seven different classes each have a SVM classification model, and then we compare the score to determine the classification result [4].

3. Developed System.

In this research, the process of facial expression recognition is divided into two parts which is the training or building the classification model and the second is testing the model which will gave the accuracies to be analysed. First, a set of training data is going through feature extraction and feature selection, the resulting data is used to build the classification model. Next is the testing process, where the testing data also goes through feature extraction and feature selection and then by using the classification model the data is classified, producing a set of predicted label. The whole process can be observed in figure 3.

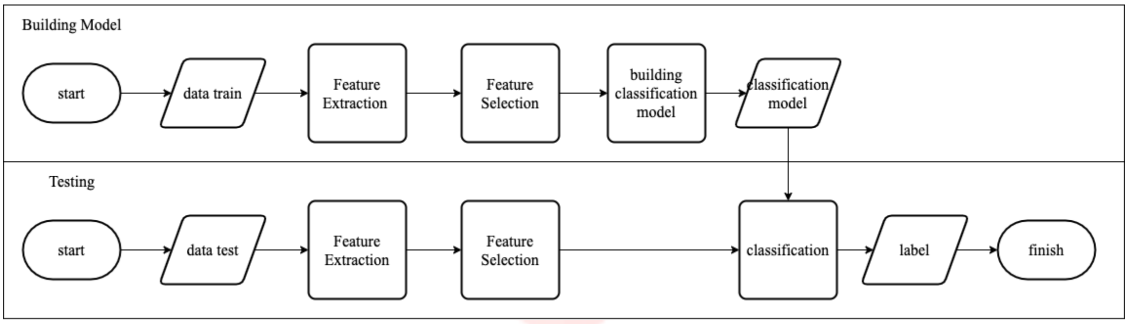


Figure 3. System flow illustration

3.1. Dataset

The dataset that is used is the JAFFE dataset or Japanese Female Facial Expression that contain labelled images of expression in each picture. The dataset has two hundred and thirteen pictures of human expression gathered from ten Japanese woman model, each with seven types of facial expression (6 basic facial expression plus 1 neutral expression) [11]. The picture itself is already in greyscale with the size of 256x256, with .tiff file type without compression. The dataset is purposed in non-commercial research and can be acquired for free in the internet. Figure 4 shows the sample of the seven-facial expression of a model from the dataset.

Figure 4. Samples of the dataset

3.2. System method flow

3.2.1. Feature Extraction

The input image that is being used to the training and testing process must first go through the feature extracting process of biorthogonal wavelet transform which is series of low pass and high pass filter. In biorthogonal wavelet, there are also sub-wavelets where each has its own kernel for the digital filter. Based on the usage, we can than consider the level of wavelet transform decomposition, if it is already satisfied the requirement then we can output the coefficient subbands as shown in figure 5.

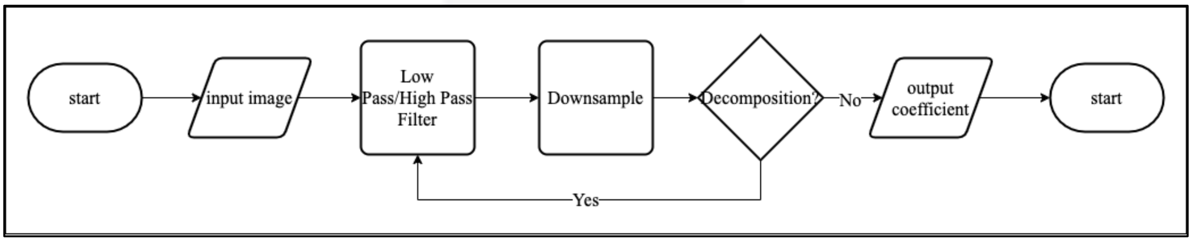
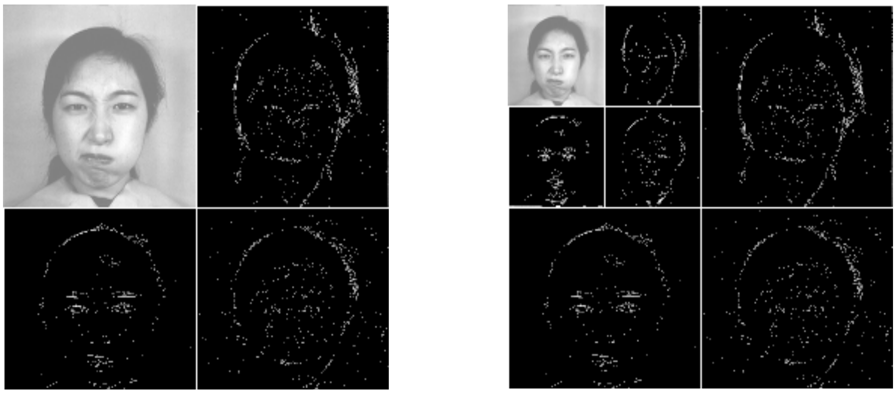


Figure 5. Feature extraction flowchart

Figure 6 presents the result of feature extraction BWT level-1 and level-2 decomposition. We can observe that the higher the decomposition level is, the more information can be retrieved.



(a) (b)

Figure 6. BWT result. (a)1-level (b)2-level

3.2.2. Dimension Reduction

The process after feature extraction is the dimension reduction, which is used for each coefficient subbands that previously produced. Using the low variance filter, areas that has low variance is filtered based on its variance and threshold. Therefore, a variance calculation needs to be done first, and to find the threshold, we start with the average variance and move forward or backward based on the classification result and the desired reduction percentage. Based on the calculation and the input threshold, a thresholding process is applied where all the element that is lower than the threshold will be deleted resulting a lower dimensional matrix from original coefficient subband matrix, the whole process can be seen in figure 7.

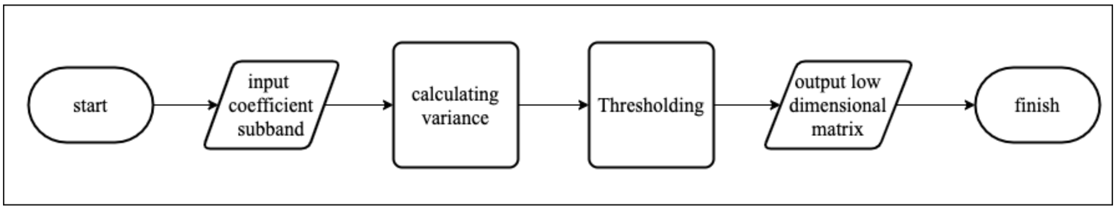
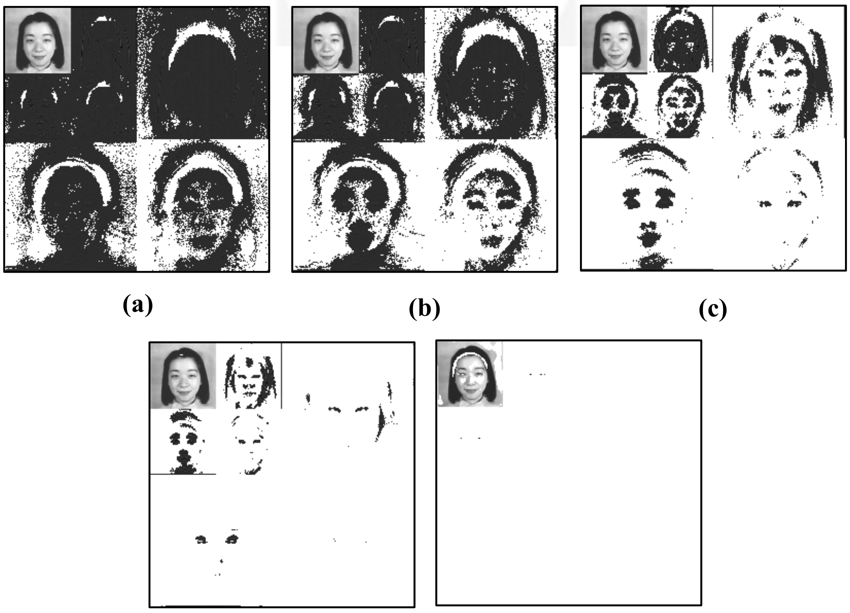


Figure 7. Data reduction flowchart

Figure 8 provides a visualization after the data is reduced, the deleted data is replaced with a white pixel for visualization purpose, while the darker ones are the data that are not deleted.



(a) (b) (c)

(d) (e)

Figure 8. Result sample from low variance filter. (a) 25% reduction. (b) 50% reduction. (c) 75% reduction. (d) 90% reduction. (e) 95% reduction

3.2.3. Classification and validation

In the classification process, the data are divided into two parts for training and testing this can be observed in figure 9. The first part is used to build the classification model, and the second part is used to test the classification model. We use the Winner-takes-all approach to deal with multiclass classification. The SVM model use the linear kernel and c=1,0.

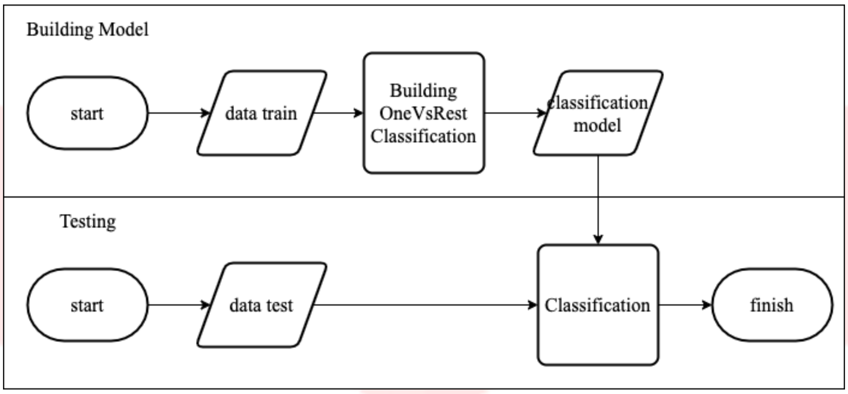


Figure 9. Classification flowchart

The 10-fold concept is to split the whole data into 10 partitions, where each is used as testing set resulting in 10 different training-testing classification models. When a partition is used, the rest of partition is used as the training set for the model. A visualized concept is shown in Figure 10. And since we use the stratified cross validation, then the train set is gathered is a manner that each fold it tries to contain the same distribution of expression input.

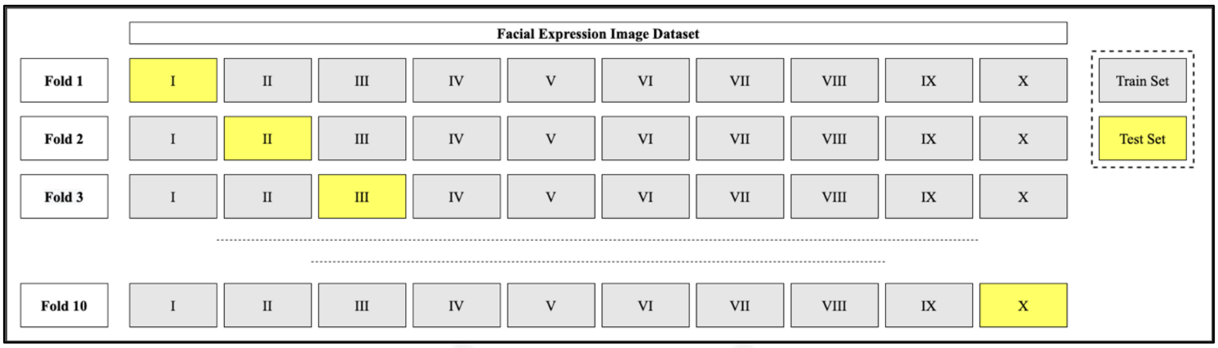


Figure 10. Data splitting map

4. Evaluation

The main purpose of this experiment is to examine how effective the purposed methods in the reduction dimension problem. It is important to find how far the dataset can be reduced and still produce the relatively the same accuracy as the one without data reduction.

There are two main part of the experiment; first, finding the best biorthogonal wavelet that produce the highest average accuracy; second, is to find how many data that can be reduced from the dataset before the accuracy is fallen. The results will be thoroughly presented in the next part.

4.1. BWT experiment result

The purpose of this scenario is to find the best biorthogonal wavelet that produce the highest accuracy between wavelets. From many wavelet, we choose six different wavelets with 2-level decomposition to extract the features and directly used to build linear SVM classification model, validating it with 10-fold cross validation. We stop at 2-level decomposition because further decomposition will decrease the performance. In the feature extraction process, BWT produces up to 65536 number of features that will be used as an input data. Bior3.1 performed poorly if compared to the other wavelet types with the overall accuracy of 89,48%. And as we can observe from figure 11 that bior4.4 bior5.5 have the same average accuracy of 93,77%.

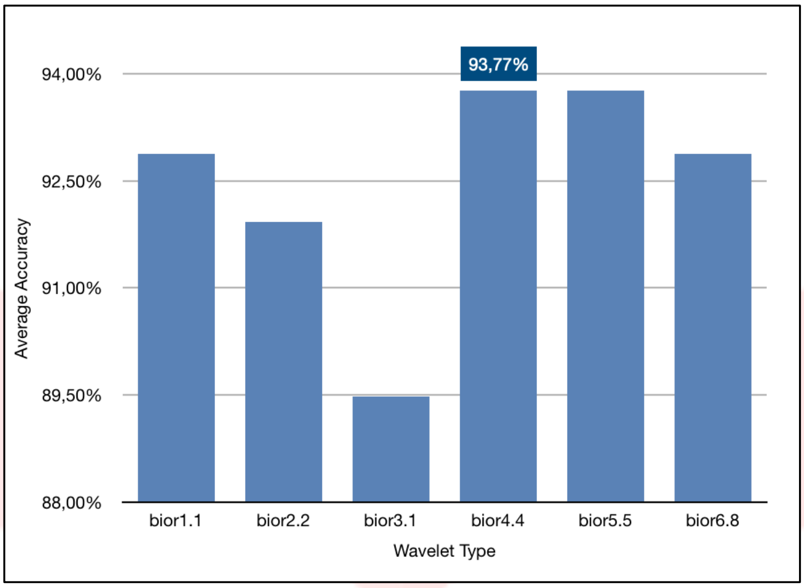


Figure 11. average 10-fold cross validation accuracy between wavelets

Since bior4.4 and bior5.5 produce the same average accuracy, a further analysis is applied. We found that the standard deviation (σ) from the two is a little different, where bior4.4 is slightly lower than bior5.5 as shown in table 2. That is why bior4.4 is considered the best wavelet compared to another wavelet.

Table 2. 10-fold cross validation comparison bior4.4 and bior5.5

	Fold-1	**Fold-2**	**Fold-3**	**Fold-4**	**Fold-5**	**Fold-6**	**Fold-7**	**Fold-8**	**Fold-9**	**Fold-10**	**σ**
Bior4.4	95,83%	100%	90,47%	100%	100%	100%	90,47%	90,47%	90,47%	80%	0,0626
Bior5.5	95,83%	100%	95,23%	100%	100%	100%	90,47%	85,71%	90,47%	80%	0,0661

4.2. BWT and low variance filter experiment result

The second scenario is focused on the data reduction. From the previous scenario, we have the best parameter for biorthogonal wavelet transform which is 2-level decomposition, and wavelet type bior4.4 is used. This experiment is searching for the maximum threshold that can be used before it affect the accuracy of the model. The results are shown in figure 12.

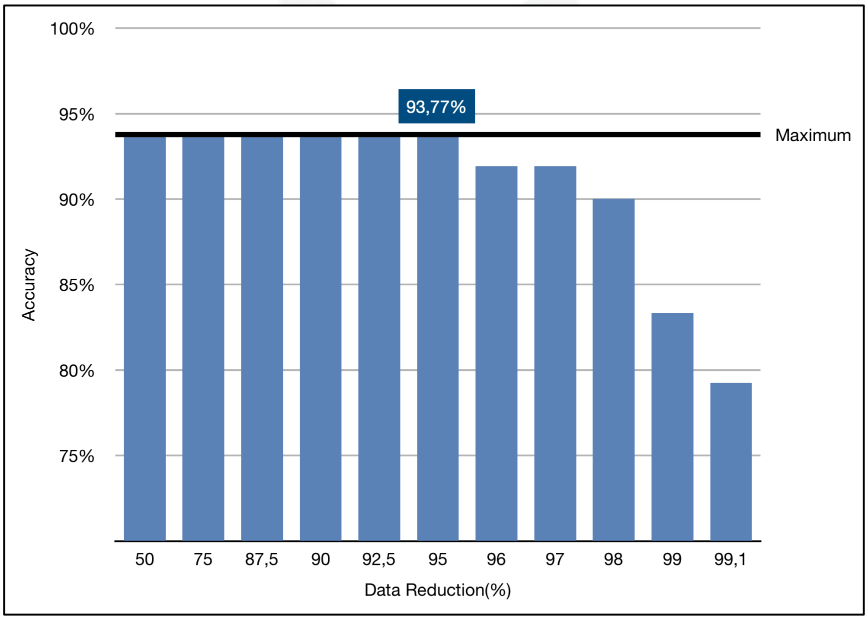


Figure 12. Average accuracy after reduction chart

As can be seen in figure 12, with the accuracy of 92% it appears that the data can be reduced up to 95% or reducing 62260 features using the threshold parameter equals 9530 before the reduction affect the accuracy.

4.3. Confusion matrix analysis

Confusion matrix will provide further information in the model performance, more precisely in terms of each class classification performance. Table 3 provide the confusion matrix before the reduction using first scenario parameter and table 4 provide the confusion matrix after dimension reduction, the row indicates the actual label and columns provide predicted label.

Table 3. confusion matrix before reduction

	HA	AN	DI	FE	NE	SA	SU
HA	26	0	0	0	3	1	1
AN	0	30	0	0	0	0	0
DI	0	2	27	0	0	0	0
FE	0	0	2	29	0	0	1
NE	0	0	0	0	30	0	0
SA	1	0	0	0	0	30	0
SU	2	0	0	0	0	0	28

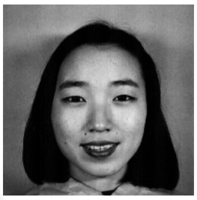
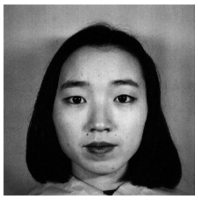
Table 4. confusion matrix after reduction

	HA	AN	DI	FE	NE	SA	SU
HA	26	0	0	0	3	1	1
AN	0	30	0	0	0	0	0
DI	0	1	27	0	1	0	0
FE	0	0	2	29	0	0	1
NE	0	0	0	0	30	0	0
SA	1	0	0	0	0	30	0
SU	1	0	0	0	0	1	28

From table 3, we learn that all Angry(AN) and Neutral(NE) expression has been correctly classified, these is caused that in AN many facial muscles are used, and in the other hand the NE used too little facial muscle; and that both proofed to be an extreme case in the dataset.

The most significant misclassified is expression Happy(HA) is confused as Neutral(NE), we found that this is happened because of Happy expression is often have the same neutral eye position with the difference in the mouth that does not shows teeth as shown in Table 13 where Euclidean distance is applied to mathematically proofed it that the input image in closer to neutral than happy interms of euclidean distance.

Table 5. Misclassification Analysis using Euclidean distance

	 Happy Expression	 Neutral Expression
 HA misclassified as NE	7385 unit of distance	6003 unit of distance

We also found that after the reduction, the misclassified images are the same from before data reduction except from one image, and only three out of 13 misclassify that have different label. A further performance comparison based on recall, precision, and f1 score [12] are provided in table 3. The same precision accuracy means that those two models make the same number of misclassification per label, but there is a slight difference in the label of misclassification resulting a little difference in recall.

Table 3. Before and after reduction model performance comparison

	Number of Data	Accuracy	Macro-Recall	Macro-Precision	Macro-F1
BWT Only	65536	93,77%	93,9322%	93,9581%	93,8531%
BWT+Low Variance Filter	3274	93,77%	94,0076%	93,9581%	93,8576%

5. Conclusion

In this study, we build a facial expression recognition model that classify seven expressions that emphasize in feature selection for dimension reduction. The methodology that are used is Biorthogonal Wavelet Transform, Low Variance Filter, and Support Vector Machine. The results of the experiment proofed that Biorthogonal Wavelet Transform is highly compatible with Low Variance Filter, producing up to 93% classification accuracy even after a 95% data reduction or reducing 62262 features and only using 3274 features from total of 65536 features by using threshold equals 9530 threshold unit with the classification model that uses the linear kernel and c equals 1.

Suggestion for future work related to the experiment is try to use other classification methods to unveil how it affect the model, also, try a more challenging dataset that not only contain female models, containing accessories, and or include image from different angels.

References

[1] Surcinelli, Paola, et al. " (Surcinelli)." *Journal of anxiety disorders* 20.1 (2006): 110-117.

[2] Zhang, Li, et al. "Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot." Expert Systems with Applications 40.13 (2013): 5160-5168.

[3] Zhang, Yu-Dong, et al. "Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation." IEEE Access 4 (2016): 8375-8385.

[4] Jabid, Taskeed, Md Hasanul Kabir, and Oksam Chae. "Robust facial expression recognition based on local directional pattern." ETRI journal 32.5 (2010): 784-794.

[5] Fabijańska, Anna. "Variance filter for edge detection and edge-based image segmentation." *Perspective Technologies and Methods in MEMS Design*. IEEE, 2011.

[6] Sifuzzaman, M., M. R. Islam, and M. Z. Ali. "Application of wavelet transform and its advantages compared to Fourier transform." (2009).

[7] Primer, A., C. Sidney Burrus, and Ramesh A. Gopinath. "Introduction to wavelets and wavelet transforms." (1998).

[8] Prasad, P. M. K., D. Y. V. Prasad, and G. Sasibhushana Rao. "Performance analysis of orthogonal and biorthogonal wavelets for edge detection of X-ray images." *Procedia Computer Science* 87 (2016): 116-121.

[9] Paul, A. "Rule-based classification of water in Landsat MSS images using the variance filter." *Photogrammetric Engineering & Remote Sensing* 63.5 (1997): 485491.

[10] Osuna, Edgar, Robert Freund, and Federico Girosit. "Training support vector machines: an application to face detection." Proceedings of IEEE computer society conference on computer vision and pattern recognition. IEEE, 1997.

[11] Lyons, Michael J., et al. "The Japanese female facial expression (JAFFE) database." *Proceedings of third international conference on automatic face and gesture recognition*. 1998.

[12] Powers, David Martin. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation." (2011).