

# Detection of Class Regularity with Support Vector Machine methods

Ni Wayan Emmy Rosiana Dewi<sup>a1</sup>, I Gede Aris Gunadi<sup>a2</sup>, Gede Indrawan<sup>a3</sup>

<sup>a</sup>Department of Computer Science, Ganesha University of Education  
Jl. Udayana No.11, Banyuasri, Kec. Buleleng, Kabupaten Buleleng, Bali, Indonesia

<sup>1</sup>[emmy.rosiana@gmail.com](mailto:emmy.rosiana@gmail.com) (Corresponding author)

<sup>2</sup>[igagunadi@gmail.com](mailto:igagunadi@gmail.com)

<sup>3</sup>[gede.indrawan@gmail.com](mailto:gede.indrawan@gmail.com)

## Abstract

*One of the most factor that affects the achievement and learning motivation of students is a conducive classroom environment. It can be seen from the student's regularity in the class. Teachers can determine whether the class is adequate or not by monitoring the class condition through video. The research tries to apply the extraction of imagery and sound features by using the Centroid extraction method and the MFCC along with classifying the regular or irregular classrooms with the SVM methods which are taken by video installed in a classroom. The video will be split into image data and sound data. The process of image data starts with reading the input, then it goes to the stages of preprocessing, segmentation with K-Means, morphology, and the most important part is to get information before it is classified by the SVM method to get its class regularity. The sound frequency will be extracted by the MFCC method and then it is classified by the SVM method to get the class noise. The results of this research get an accuracy value of 78% in the linear kernel and 70% in the polynomial kernel. This research uses 50 test data consisting of 25 regular data and 25 irregular data taken directly through video recording. These results prove that the SVM method has given good classification results for regular and irregular classes.*

**Keywords:** Class, Centroid, MFCC, SVM, Regular, Irregular

## 1. Introduction

A classroom is a place for intensive teaching and learning activities. Students and teachers interact, give, and receive lessons in class, to achieve the objectives of national education. One of the factors that influence student achievement and motivation is a conducive classroom environment. A motivating environment will make it easier for students to accept lessons, in addition to be able to develop initiatives (the desire to learn on their own). Achievement of student learning achievement can be improved by evaluating the conditions of student learning activities through video recording media.

Monitoring the condition of the classroom through video is also one application of the video that helps the teacher to review whether the class is conducive or not. Image and audio data is captured via video, where each information is processed based on its features. The level of regularity and class noise can affect student's motivation and learning achievement [1]. Therefore in this study, it is trying to apply image and sound feature extraction by using the Centroid Extraction and Mel Frequency Cepstral Coefficient (MFCC) method and classifying regular or irregular classrooms with the Support Vector Machine (SVM) method which is taken through video attached in a class. This research is a basic research that can be used to build a system which is called integrated smart class. One of them contains a feature of monitoring classroom conditions in real/through video (images and sounds). It aims to make it easier for teachers to monitor when the teacher isn't there in the class or the student studies independently.

Several previous studies related to the classification method using the Support Vector Machine (SVM) have been done. One of these was by I Gede Aris Gunadi and friends, with the title Fake Smile Detection Using Linear Support Vector Machine [2]. In this study, the detection of smiles from people's faces, whether smiles are real or fake by using the RoI (Region of Interest) segmentation technique, was done on the cheeks and eyes. The test results show that the accuracy of the system is 86%, while the error rate is 14%. Other research on SVM classification is a study conducted by Raudlatul Munawarah and friends with the title "Application of the Support Vector Machine Method in Hepatitis Diagnosis [3]. This study analyzes the ability of the SVM method to use training data of 100 positive and 100 negative data using linear kernel functions and RBF.8 The percentage results of the classification using linear kernels are 68-83% and kernel RBF by 70-96%.

Research on the image of the classroom environment was carried out by researchers Takashi Ozeki and Watanabe, who made a study entitled Analysis of the Behavior of Students Considering Privacy [4]. This study uses the Haar classifier method for smoothed video. Then, check the pixel number of the skin color of the face area detected by this method, then each face is given a number. From the experiments, it was possible to determine the classification correctly when students faced forward even in smoothed videos. Research on image feature extraction has been carried out by Kadek Novar Setiawan and friends using the K-Means GLCM method in obtaining image feature extraction. The application of the k-means method is used in the segmentation process with 4 clusters. The GLCM method is used in the image extraction process, which aims to extract relevant information into the characteristics of each class. Support Vector Machine used as a classification process shows good results in distinguishing normal and abnormal mammogram images by showing an accuracy of up to 80%, so this method is considered good enough to be used in the classification process of mammogram images [5].

The research about extraction of sound features using MFCC has been researched by Awais, et al. Their research was using MFCC as extraction feature methods of speech signal with locality sensitive hashing (Lsh) as its clarification method. The research got 92.66% accuracy values for the speech recognition process by matching the training data that it has [6]. Other research on sound or audio extraction using the MFCC method was conducted by Mohan B and Ramesh Babu N with the title Speech recognition using MFCC and DTW research. This study extracts sound features using the Mel Frequency Cepstral Coefficients (MFCC) and Dynamic Time Wrapping (DTW) methods, two algorithms, each of which is adapted for feature extraction and pattern matching. Results obtained with one training and continuous testing phase [7].

Based on these studies, research about the detection system for the class regularity using the image and sound features with the Support Vector Machine (SVM) method has not been done yet. SVM is a machine learning method that is supervised learning which is still relied on in terms of binary classification and while this SVM method is not used yet in classifying object images and sounds in a classroom together. These two characteristics data are then modeled by the SVM Method as Training and Classifying, whether class conditions are regular or irregular. So it is hoped that This research can contribute in the form of class image datasets and audio class datasets because in the process of data acquisition, it is done directly by using the same data collection standards both from the tool and the angle of video data retrieval, which is separated into images and audio, and Hopefully this research can contribute references for other research in knowing the image and audio classification process by using the SVM method.

## 2. Research Methods

This study took data directly from the condition of the classroom during class hours. Sibangkaja Public Elementary School 4 was a place for collecting data used in this study. The image of this class condition taken using a Fujifilm X-T100 Mirroring Camera recording device; then, the recorded file was stored and then processed with video processing software to take pictures and audio for 5 seconds. Image data in JPEG format (24-bit color depth), resolution of 1980 x 1080 pixels with the highest quality 96 dpi. Audio data then saved in .wav format.

An overview of this research presented in the following Figure 1.

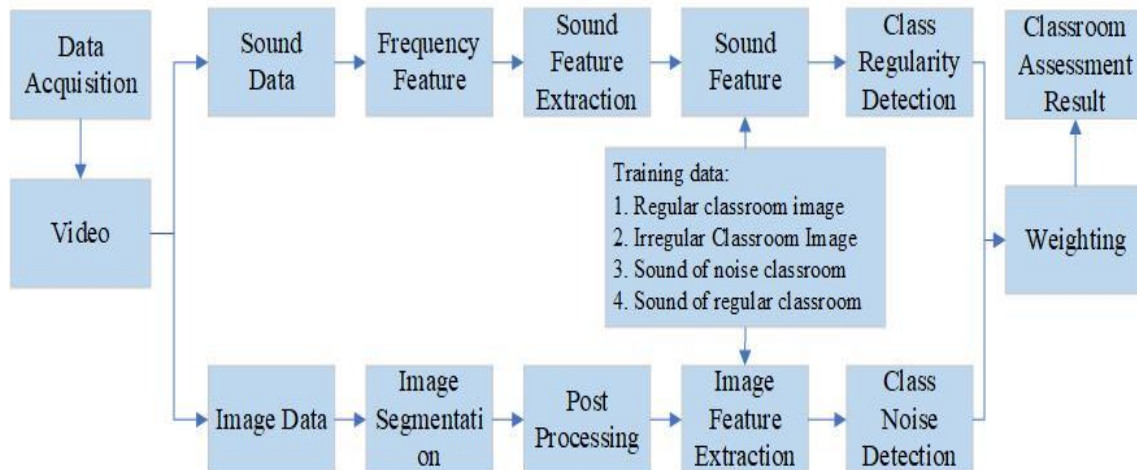


Figure 1 Overview of the method approach for class regularity detection

Class regularity detection uses two inputs derived from images and audio. Each input must produce features that can be used by SVM classification methods to determine the conditions of regular or irregular classes. For example, for the image input, the hair position feature used as a feature, and the audio input use the value of the intensity of the sound frequency produced by students in the class as a feature.

The hair position feature is used as a feature value, assuming that if students pay attention in the class, they will regularly sit with the hair position will look regular if drawn in a straight line horizontally. Conversely, if students who are in the class do not focus ahead, of course, the position of each student's hair will look irregular. Characteristics of hair position in the study using the centroid value of each segmentation obtained. As for audio input, the characteristic value is taken from the intensity value of the sound frequency produced. Assume used is the higher the intensity of the sound frequency obtained from the input, the class tends to be irregular, and conversely, the lower the intensity of the sound frequency obtained, the condition of the class tends to be regular.

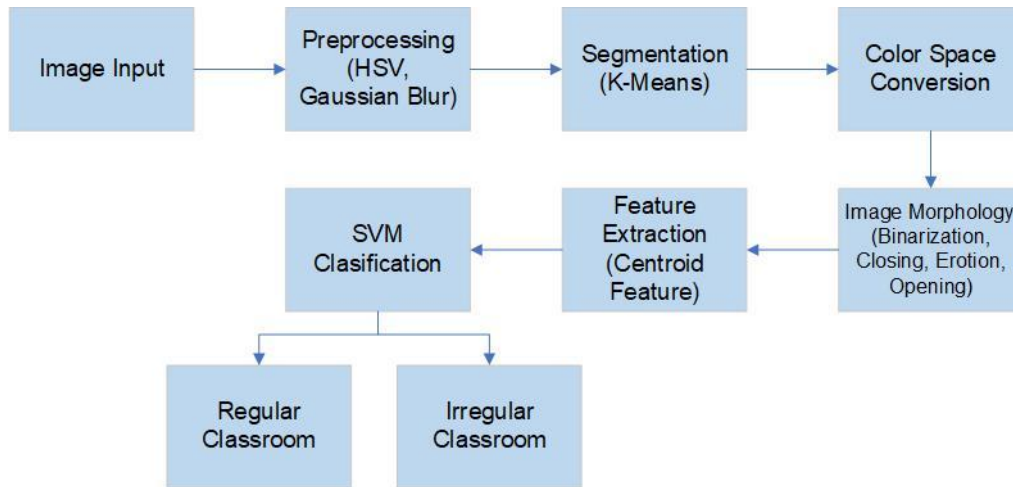
The detailed process of each input, image, and audio can be seen as follows.

## 2.1. Image Data

Image data is an image that is similar to its original form or at a minimum in the form of a planimetry. Images or digital images on a two-dimensional scale are processed and manipulated by the image processing method [8]. The image processing process in this study is seen in Figure 2. In the preprocessing process, the input image in the RGB color space is converted to HSV. This color model is in accordance with human perception of the similarity of colors [9]. The Gaussian blur filtering process is included in the preprocessing image, where the image is blurred and reduced the noise contained in it [10]. The next stage is image segmentation using K-Means. This study will look for students' hair objects using K-Means segmentation. Segmentation is a technique for dividing an image into several regions where each region has a similar attribute [11][12]. K-Means is an unsupervised clustering algorithm and it is used to segment more prominent areas of the background [13]. K-Means can work well on image segmentation, if the image has previously been partially repaired [13]. Furthermore, the segmented image will be processed by image morphology into several steps. First is binrization, which changing the image in binary form, namely an image with two gray level values, black and whites[14]. Then, closing which smoothing the segmentation and cover the missing pixels . The last one is erosion, which moving pixels at object boundaries and opening refine object boundaries, separate objects that were previously hand in hand, and eliminate objects smaller than the size of the structuring [15][16].

The process after image morphology is feature extraction by finding the centroid position of each student's hair in the class. The regular position of the hair centroids in each image

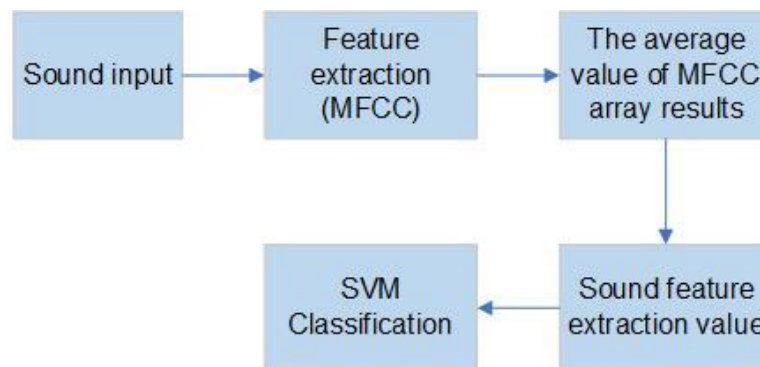
determines the regularity of the class. So when another image is tested, the Support Vector Machine algorithm will classify whether the image is organized or not.



**Figure 2** Image data processing

Feature extraction is an essential step in the decision-making process in determining whether the object is in an organized position or not based on the student's position. This feature is also for determining unknown objects in the class. Image data used in this research use the .jpg format. Training data used in this research were 125 data, consisting of 63 regular student data and 62 irregular student data

## 2.2. Audio Data



**Figure 3** Audio data processing

Audio data processing begins with feature extraction, which in this stage, a series of quantities in the input signal section are processed to determine learning patterns or test patterns. The features used in this study are frequency features. For sound signals, the magnitude characteristic is usually the output of some form of spectrum analysis technique, which in this study uses the MFCC (Mel-Frequency Cepstrum Coefficients) method. MFCC is a feature extraction that calculates the cepstral coefficient by considering human hearing [17]. MFCC values used in this study were 20 values from 0-19. The audio format used in this study is .wav.

## 2.3. Support Vector Machine (SVM)

The Support Vector Machine (SVM) developed by Boser, Guyon, Vapnik, and was first presented in 1992 at the Annual Workshop on Computational Learning Theory. The basic concept of SVM is data calculation techniques. By using statistics and learning with expected

results in the form of predictive abilities. SVM can be applied to results which is continuous, binary, categorical, logistic, and multinomial by forming a hyperplane margin [18][19]. SVM uses kernel assistance to connect training data input to wider space dimension features and identifies its hyperplane as a dividing space [20].

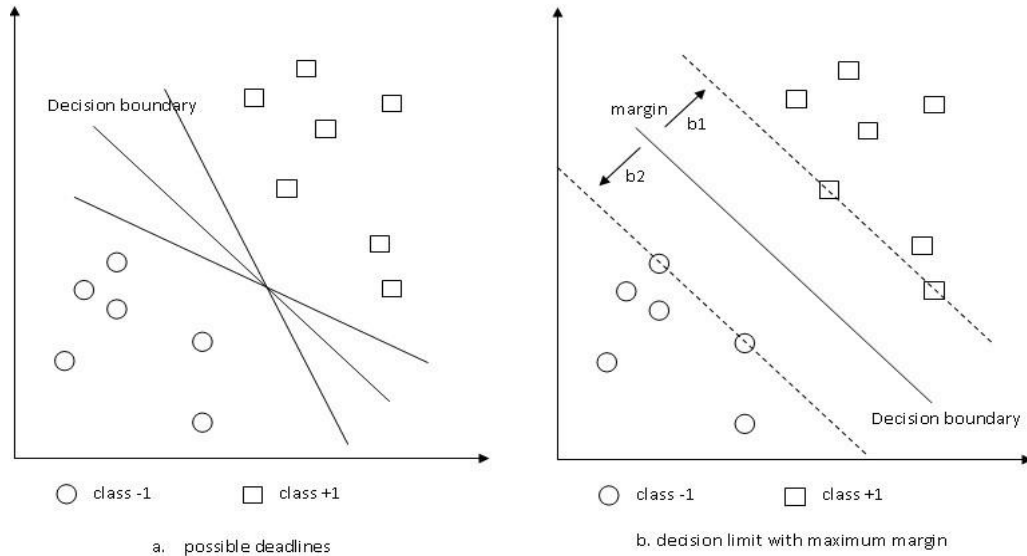


Figure 4 SVM visualization

The concept of classification with SVM can be explained simply as an attempt to find the best hyperplane that functions as a separator of a two-class or multi-class in the input space [21]. Figure 4 shows some data that are members of two data class pieces, namely +1 and -1. Data incorporated in class -1 is symbolized by a circle, while data in class +1 symbolized by a square [22].

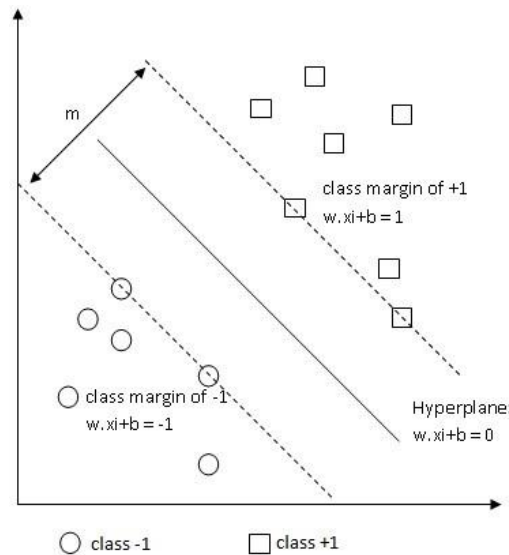


Figure 5 Hyperplane SVM margin

The best separator hyperplane (decision boundary) between the two classes can be found by measuring the margins and finding the maximum point. Margin is the distance between the hyperplane and the closest data from each class. The closest data is referred to as a support vector. The solid line in Figure 5 to the right shows the best hyperplane, which is located right in the middle of the two classes, while the circle and square data that is crossed by the margin line

(dashed line) is the support-vector. Efforts to find the location of this hyperplane are the core of the training process in SVM.

### 3. Result and Discussion

The algorithm proposed in this study was created using the Python programming language. The training data used were 125 data obtained through direct recordings from two different classrooms. The training process is ready after the image pre-processing and feature extraction process is complete.

The test carried out using two existing kernels in SVM, namely a linear kernel and a polynomial kernel. The type of kernel is the parameter used to modify the best separator hyperplane in the SVM input space [23]. Choosing the right kernel function is very important because this kernel function will determine the feature space where the classifier function will be searched for. As long as the kernel function is legitimate, SVM would operate correctly, even though we didn't know what map to use [24]. In the next step, SVM would use hyperplane as a decision boundary efficiently.

#### 1. Linear Kernel

The linear kernel is the most straightforward kernel function. It is used when the data analyzed is linearly separated. Linear kernels are suitable when there are many features because mapping to higher dimensional spaces cannot improve performance as in text classification. In-text classification, both the number of instances (documents) and the number of features (words) are the same. The following is the equation of the SVM linear kernel.

$$K(x, x') = x \cdot x' \dots\dots\dots (1)$$

where  $x$  and  $x'$  are vectors in the input space.

#### 2. Polynomial Kernel

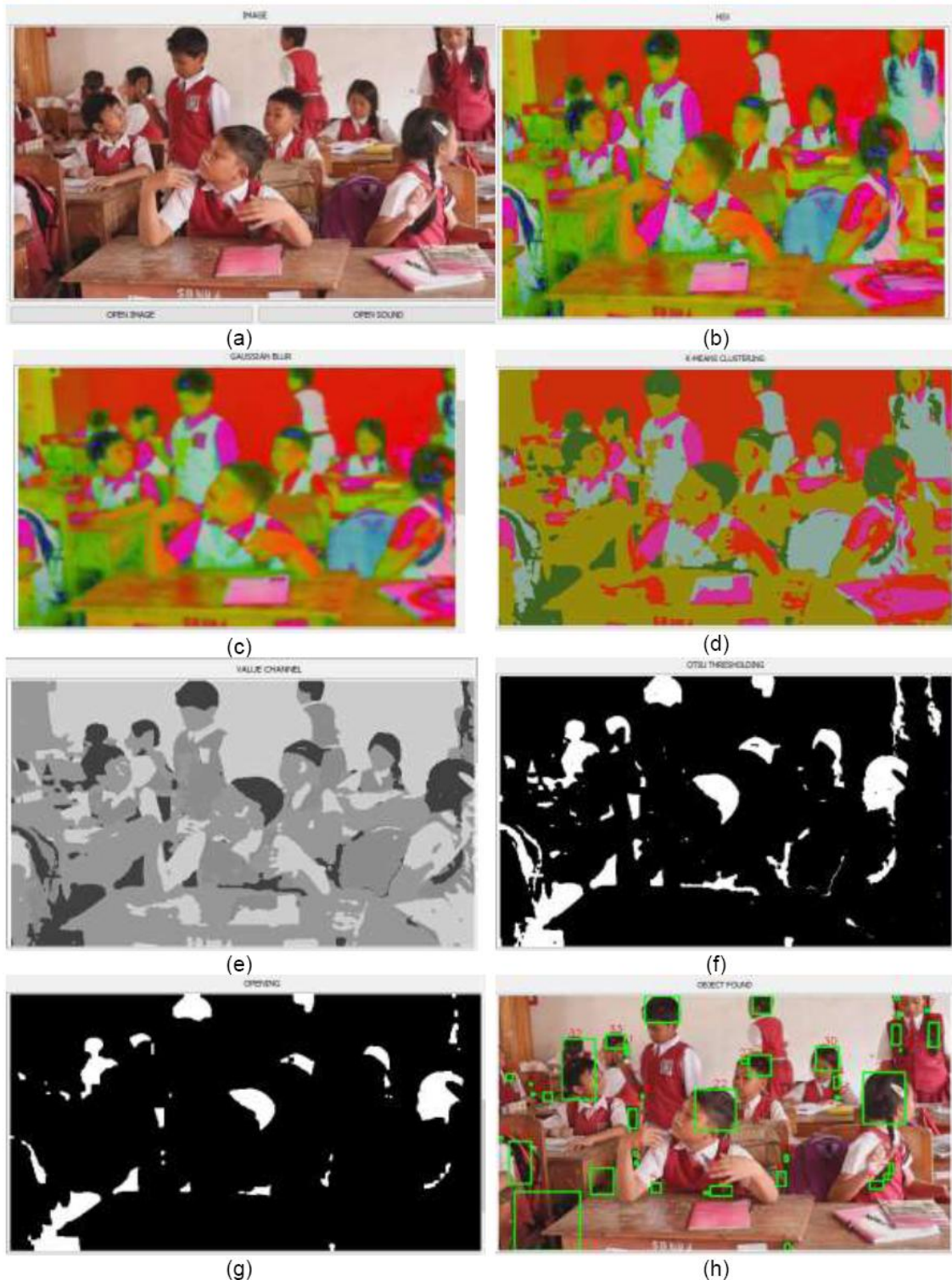
The kernel polynomial is a kernel function that is used when data is not linearly separated. The polynomial kernel is perfect for problems where all training datasets are normalized, along with the polynomial equation.

$$K(x, x') = (x \cdot x' + c)^d \dots\dots\dots (2)$$

It has two parameters:  $c$ , which represents a constant term, and  $d$ , which represents the degree of the kernel.

### 3.1. Training Data

The training process on this system begins by entering all the image and audio data that has been prepared as training data. A total of 125 data are used as training data. After the data is inputted, proceed with the preparation process. Trained data is displayed in sequence, starting from the results of image preprocessing, which consists of the conversion of RGB images into HSV images, image filtering using the Gaussian Blur method, and image segmentation using the K-Means method. Parameters used in segmentation with k-means clustering are  $K = 5$  and 10 iterations. After segmentation with k-means, the Hue, Saturation, and Value channel channels are separated. Post-processing consists of Otsu Thresholding, closing, erosion, and opening after the V channel has been determined. The image that has been through post-processing is then carried out the process of extracting image features through centroid extraction by finding the coordinates of the center point  $(x, y)$  of each object. The stages of the training data image processing can be seen in Figure 6 below.



**Figure 6** The image processing process to get the centroid characteristics that will determine the coordinates of students' hair (a. Image input, b. Image of HSV results, c. Image of Gaussian Blur filtering result, d. Image of K-Means segmentation result, e. Image of V Channel results, f. Image of Otsu Thresholding results, g. Image of Opening results, h. Image of Centroid Feature Extraction)

Audio data processing using MFCC is displayed in tabular and graphical form, as shown below. Voice feature extraction with MFCC produces an array of MFCC results of 20 values, which are then used as a feature value in the process of establishing the model during training.

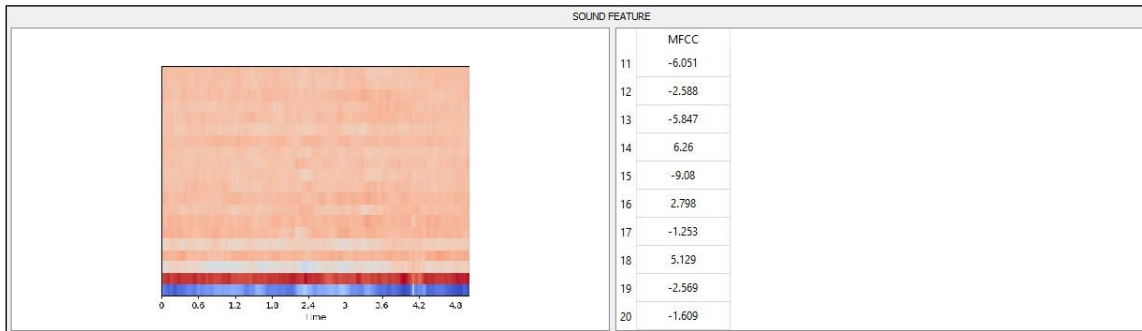


Figure 7 Audio processing with MFCC

The graph in Figure 7 shows the spectrum of MFCC values generated within 5 seconds while the MFCC values in the table are presented as many as 20 values. Cepstrum, in the form of the coefficient value of the features/features of the sound signal is the result of the MFCC feature extraction method, which is to get the coefficient value as the typical value of the sound signal so that the sound signal pattern is easily recognized. The process of modeling data in the training menu is done after all training data has been entered.

### 3.2. Testing and Evaluation Data

The results of testing the data performed by the system can be seen in Figure 8. Test data that has been prepared through the acquisition phase will be processed to produce a classification of data which is then stored after going through an evaluation process by an expert. Tests carried out using the same 50 test data in each kernel. The use of kernels in SVM aims to classify data that cannot be classified linearly. SVM is the most well-known method with a wide range of data classes that uses the kernel to represent data and can be called a kernel-based method [25].

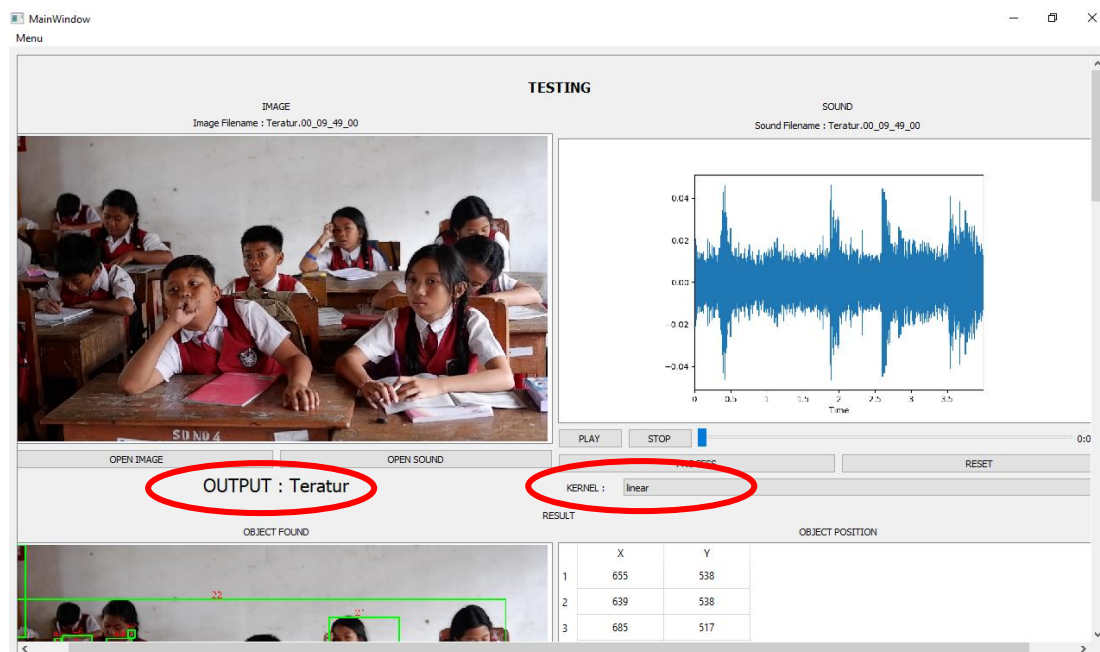


Figure 8 Testing data interface





Figure 9 Evaluation result interface

After testing the data, the result will be evaluated by experts in this case conducted by the teacher to see the comparison of the results of the classification carried out by the system with actual conditions. The evaluation menu interface shown in Figure 9. The test results for each kernel are presented in the confusion matrix as follows.

### 3.2.1. Linear Kernel

Table 1 Confusion matrix table in a Linear kernel with 50 test data

N = 50	Actual	
	Regular	Irregular
Predict	Regular	8
	Irregular	16

Based on Table 1 above, the calculation results obtained are 78% accuracy, 74% precision, 89% recall, and F-measure value of 80% for testing in linear kernels.

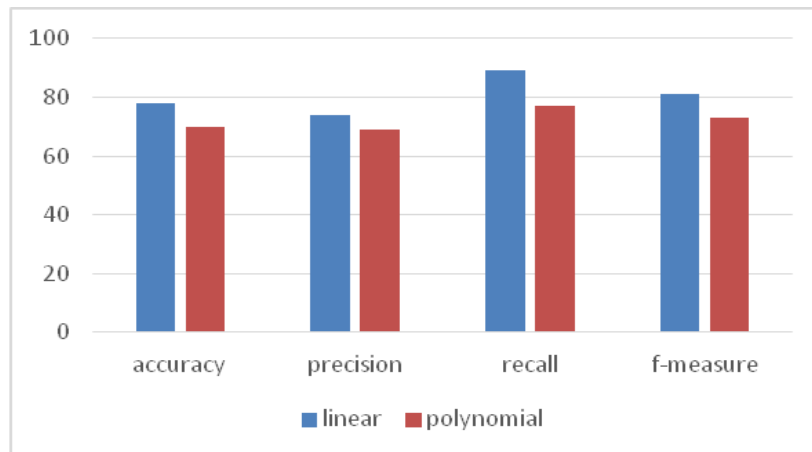
### 3.2.2. Polynomial Kernel

Table 2 Confusion matrix table in a Polynomial kernel with 50 test data

n = 50	Actual	
	Regular	Irregular
Predict	Regular	9
	Irregular	15

Based on Table 2 above, the results obtained are 70% accuracy calculation, 69% precision, 77% recall, and F-measure value of 73% for testing on the polynomial kernel.

A comparison of the results of accuracy, precision, recall, and f-measure linear and polynomial kernels can be seen more clearly in the following graph.



**Figure 10** Comparison graph of the accuracy, precision, recall, and f-measure of linear and polynomial kernels

Figure 10 shows that the linear kernel produces an average success rate in classifying regular and irregular classes higher than the polynomial kernel, seen from the level of accuracy, precision, recall, and f-measure. Linear kernels detect true data more than actual polynomial kernels by using the same 50 test data for each kernel because the linear kernel separates the data linearly and straight line. The same results are obtained by Supriya Pahwa with the research entitled “Comparison Of Various Kernels Of Support Vector Machine” which in his research stated that linear kernel gives the best performance an average of 88.20% correct classification compared to other types of kernel functions [26].

#### 4. Conclusion

This research aims to classify the condition of the classroom whether regularly or irregularly as we can see problems that occur when the teacher is not in class, students tend to make noise. Based on experiments which were conducted in this research, the number of conclusions can be drawn, that to obtain the information whether the class is regular or not, the image and audio data of the class conditions must go through a processing stage first. The image was processed through the stages of preprocessing, segmentation with K-Means and hair centroid extractions, which were used as features in this study. The method used for sound feature extraction in this research is MFCC. The test was carried out by using 125 training data and 50 data for each kernel, it obtained accuracy on the linear kernel of 78% and 70% polynomial kernel. It can be concluded that SVM works well in linear kernels in classifying regular and irregular classes.

#### References

- [1] S. Suprihatin, “Upaya Guru Dalam Meningkatkan Motivasi Belajar Siswa,” *PROMOSI (Jurnal Program Studi Pendidikan Ekonomi)*, vol. 3, no. 1, pp. 73–82, May 2015.
- [2] I. Gede Aris Gunadi, A. Harjoko, R. Wardoyo, and N. Ramdhani, “Fake Smile Detection Using Linear Support Vector Machine,” in *Proceedings of 2015 International Conference on Data and Software Engineering, ICODSE 2015*, pp. 103–107, 2016.
- [3] R. Munawarah, O. Soesanto, and M. R. Faisal, “Penerapan Metode Support Vector Machine Pada Diagnosa Hepatitis,” *Kumpulan Jurnal Ilmu Komputer (KLIK)*, vol. 04, no. 01, pp. 73–82, Feb 2016.
- [4] T. Ozeki and E. Watanabe, “Analysis of the Behavior of Students Considering Privacy,” in *The 6th IIEEJ International Conference on Image Electronics and Visual Computing*, no. 1P-3, 2019.
- [5] I. M. S. P. Kadek Novar Setiawan, “Klasifikasi Citra Mammogram Menggunakan Metode K-Means, GLCM, dan Support Vector Machine (SVM),” *Jurnal Ilmiah Merpati (Menara Penelitian Akademika Teknologi Informasi)*, vol. 6, no. 1, pp. 13–24, 2018.

- [6] A. Awais, S. Kun, Y. Yu, S. Hayat, A. Ahmed, and T. Tu, "Speaker Recognition Using Mel Frequency Cepstral Coefficient and Locality Sensitive Hashing," in *2018 International Conference on Artificial Intelligence and Big Data, ICAIBD 2018*, pp. 271–276, 2018.
- [7] B. J. Mohan and N. Ramesh Babu, "Speech Recognition Using MFCC and DTW," in *2014 International Conference on Advances in Electrical Engineering, ICAEE 2014*, pp.1-4, 2014.
- [8] O. Lézoray and L. Grady, *Image Processing and Analysis with Graphs: Theory and Practice*. CRC Press, 2012.
- [9] M. Loesdau, S. Chabrier, and A. Gabillon, "Hue and saturation in the RGB color space," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 203–212, Springer International Publishing, 2014.
- [10] E. S. Gedraite and M. Hadad, "Investigation on the effect of a Gaussian Blur in image filtering and segmentation," in *Proceedings Elmar - International Symposium Electronics in Marine*, pp. 393–396, 2011.
- [11] Darma Putra, *Pengolahan Citra Digital*. Yogyakarta: Penerbit Andi, 2010.
- [12] S. S. Dhumal and S. S. Agrawal, "MRI Classification and Segmentation of Cervical Cancer to Find the Area of Tumor," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 3, no. VII, pp. 21–26, 2015.
- [13] A. Mohd, G. K. Ram, and A. Shafeeq, "Skin Cancer Classification Using K-Means Clustering," *International Journal of Technical Research and Applications*, vol. 5, no. 1, pp. 62–65, 2017.
- [14] H. Kim, E. Ahn, S. Cho, M. Shin, and S. H. Sim, "Comparative Analysis of Image Binarization Methods for Crack Identification in Concrete Structures," *Cement and Concrete Research*, vol. 99, pp. 53–61, Sep. 2017.
- [15] L. Najman, J. C. Pesquet, and H. Talbot, "When Convex Analysis Meets Mathematical Morphology on Graphs," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9082, pp. 473–484, 2015.
- [16] Y. Chugh, R. Gupta, and R. Kaushik, "Image Enhancement Using Morphological Operators," *International Journal of Engineering Technology*, vol. 3, special Issue, pp. 61–66, 2015.
- [17] T. Chamidy, "Metode Mel Frequency Cepstral Coefficients (MFCC) pada Klasifikasi Hidden Markov Model (HMM) untuk Kata Arabic pada Penutur Indonesia," *Jurnal Matics*, vol. 8, no. 1, pp. 33-40, 2016.
- [18] N. Guenther and M. Schonlau, "Support Vector Machines," *The Stata Journal: Promoting Communications on Statistics and Stata*, vol. 16, no. 4, pp. 119-127, 2016.
- [19] M. Aykanat, Ö. Kılıç, B. Kurt, and S. Saryal, "Classification of Lung Sounds Using Convolutional Neural Networks," *Eurasip Journal on Image and Video Processing*, no. 65, 2017.
- [20] A. F. Indriani and M. A. Muslim, "SVM Optimization Based on PSO and AdaBoost to Increasing Accuracy of CKD Diagnosis," *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, vol. 10, no. 2, pp. 119-127, Aug 2019.
- [21] Y. R. Nugraha, A. P. Wibawa, and I. A. E. Zaeni, "Particle Swarm Optimization-Support Vector Machine (PSO-SVM) Algorithm for Journal Rank Classification," in *Proceedings - 2019 2nd International Conference of Computer and Informatics Engineering: Artificial Intelligence Roles in Industrial Revolution 4.0, IC2IE 2019*, 2019, pp. 69–73.
- [22] P. Rebentrost, M. Mohseni, and S. Lloyd, "Quantum Support Vector Machine for Big Data Classification," *Physical Review Letters*, vol. 113, no. 13, pp. 130503, Sep. 2014.
- [23] D. P. Kaucha, P. W. C. Prasad, A. Alsadoon, A. Elchouemi, and S. Sreedharan, "Early

- Detection of Lung Cancer using SVM Classifier in Biomedical Image Processing,"in *IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPSI)*, pp. 3143–3148, 2017.
- [24] R. Fernandes de Mello, M. Antonelli Ponti, R. Fernandes de Mello, and M. Antonelli Ponti, "Introduction to Support Vector Machines," in *Machine Learning*, 2018.
- [25] M. Gönen and E. Alpaydin, "Multiple Kernel Learning Algorithms," *Journal of Machine Learning Research*, vol. 12. pp. 2211–2268, Jul-2011.
- [26] S. Pahwa and D. Sinwar, "Comparison Of Various Kernels Of Support Vector Machine," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 3, no. VII, pp. 532–536, 2015.