



Linear Kernel and Polynomial Analysis in Recognizing Tuberculosis Image Using HOG Feature Extraction

Ira Farendasudirman¹, Winda Hartati Giawa², Intan Permatasari Sarumaha³, Sukurman Ndraha⁴, Insidini Fawwaz⁵

¹²³⁴⁵Teknik Informatika

¹²³⁴⁵Teknologi dan Ilmu Komputer, Universitas Prima Indonesia, Jl. Sekip simp. Sikambing, Kecamatan Medan Petisah, Kota Medan, Indonesia

E-mail: irafarendas@gmail.com

ARTICLE INFO

Article history:

Received: 12/07/2020

Revised: 22/08/2020

Accepted: 30/09/2020

Keywords:

Tuberculosis, HOG, Linear, Polynomial, SVM

ABSTRACT

Tuberculosis (TB) is an airborne disease caused by mycobacterium tuberculosis (MTB) which usually attacks the lungs which can cause severe coughing, fever and chest pain. The recognition of TB negative and positive TB x-ray image patterns in this study uses HOG feature extraction and the SVM method as a classification method by adding linear and polynomial kernel functions to the SVM method. This is because even though it is very good at solving classification problems, SVM can only be used on linear data, so that in order to be used on non-linear data, SVM must be modified using kernel functions. The results showed that the linear kernel was better at classifying the x-ray image of TB with an average accuracy of 79.50% while the polynomial kernel was 77.50%.

Copyright © 2020 Jurnal Mantik.
All rights reserved.

1. Introduction

In the health sector, the development of medical science has progressed very rapidly, marked by the discovery of new diseases that have not been previously identified. One of the diseases currently developing is tuberculosis (TB) [1]. According to [2] Tuberculosis (TB) is an airborne disease caused by Mycobacterium tuberculosis (MTB) which usually attacks the lungs which cause severe coughing, fever, and chest pain. tuberculosis (TB) is still one of the biggest causes of death in Indonesia, so it needs special handling in the prevention of tuberculosis (TB). In 2004, WHO Global Surveillance estimated that in Indonesia there were 583,000 new tuberculosis sufferers each year with 262,000 smear-positive or an incidence rate of approximately 130 per 100,000 population and deaths from tuberculosis are estimated to affect 140,000 people each year [3].

Along with the development of the times, the need for intelligent machines that can identify diseases from an image is needed, because it is very helpful in many aspects, especially in the efficiency of diagnosis in the health sector [4]. Machine learning has been widely used in the medical field to analyze medical datasets. One of the most commonly used diagnostic techniques is through X-ray images of the lungs. Every lung disease has different characteristics and patterns, as well as tuberculosis. So far, the diagnosis is still very dependent on the subject making the diagnosis, so it seems subjective and takes a long time to determine the diagnosis. One of the most frequently used machine learning methods is Support Vector Machine (SVM). The characteristic of this method is to find the optimal classifier function that can separate two data sets from two different classes so that the computation process is more efficient and does not require large memory space in the classification process [1].

Several studies that have been carried out regarding the classification of tuberculosis using machine learning, namely: [3] in his study analyzed the classification of cases of infected or uninfected tuberculosis patients using the Supports Vector Machine (SVM) and Logistic Regression (RBF) approach. The results show that SVM can measure classification accuracy with higher accuracy. The accuracy results obtained by SVM with the RBF kernel function are 98%. [5] using the SVM and CNN methods in detecting tuberculosis the results showed an increase in classification accuracy and were able to train custom architectures from scratch and achieve good results and increase accuracy with large numbers of datasets. [6] used CNN in the AlexNet, VGG-16 and CapsNet classification of pulmonary tuberculosis.

Based on previous studies, the detection of tuberculosis was carried out in several stages, including a thorough examination of lung X-ray images. However, it turns out that there are still some difficulties in



detecting tuberculosis, especially in analyzing the X-ray image of the lungs. The analysis carried out so far can be said to be subjective, because it still relies on the subject who assesses it, in this case, a medical expert. From these problems, it is necessary to have a computational-based analysis system, so that the results obtained are more accurate and objective. In this research, the HOG feature extraction and SVM method will be used with Linear and Polynomial kernel functions in determining the negative and positive results of tuberculosis (TB).

2. Research Methodology

2.1 Dataset

The data used in this study are images with a size of 256x256. Image files used for training and testing were taken from http://openi.nlm.nih.gov/imgs/collections/ChinaSet_AllFiles.zip with a total sampling of 622 tuberculosis (TB) disease images.

2.2 Research Steps

The general research steps that were built in this study can be seen in Figure 1 below:

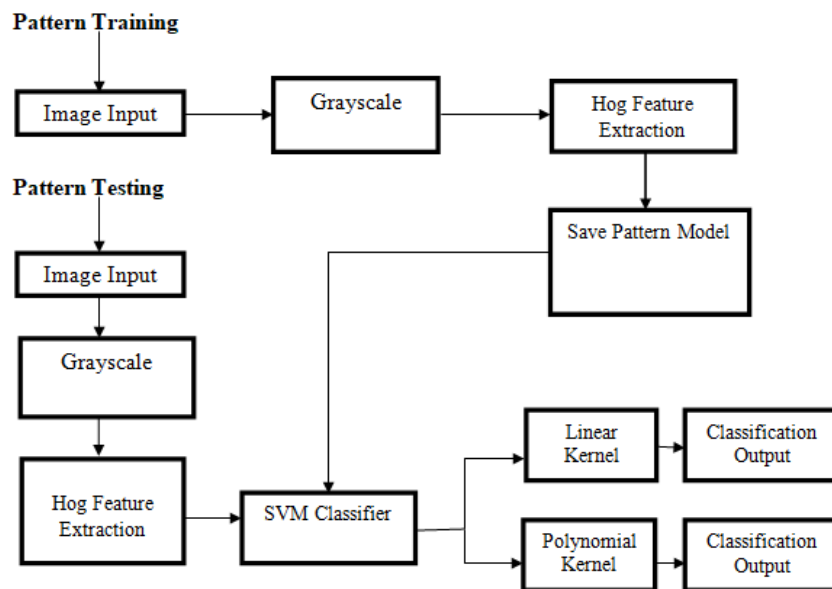


Fig 1. General Research Steps

Figure 1. illustrates a general research scheme, in this scheme there are two processes, namely the training process and the testing process, the training process for color input images will be converted into gray images to save computation time from three color channels to 1 color channel, then extracted using The HOG and the extracted results will be saved as a pattern model. While the testing process goes through the same stages as the stages that are passed during the training, only when the classification stage uses two kernels, namely: the linear kernel and the polynomial kernel on the SVM to get the best results in the introduction of tuberculosis (TB).

2.3 Histogram of Oriented Gradient

Histogram of Oriented Gradient (HOG) is a method used in image processing for object detection purposes. This technique is also commonly used in Computer Vision. To obtain differentiating information, the image will be divided into cells and each cell will be counted as a histogram of oriented gradients [8]. Each cell is compiled a histogram of a gradient. The combination of this histogram is used as a descriptor that represents an object. Each pixel in a cell has its own histogram value based on the value generated in the gradient calculation. Cells have a size of 4x4 pixels in an image while blocks are 2x2 cells or 8x8 pixels in size. The result of the HOG feature is converted into a feature vector for processing into the SVM classifier. The SVM classifier is used to find the dividing line (hyperplane) from class +1 which contains positive feature vectors to class -1 which contains negative feature vectors [9]. After being trained, the training result model is tested to determine the level of accuracy in recognizing the object being tested.

2.4 Support Vector Machine (SVM)

Support Vector Machine can be explained as an attempt to find the most optimal (best) hyperlane or boundary line that functions as a differentiator for two classes in the input space. The basic principle of SVM is a linear classifier, and was further developed so that it can work on non-linear problems. by incorporating the kernel concept into a high-dimensional workspace [3]. This development has stimulated research interest in the field of pattern recognition to investigate the potential of SVM capabilities theoretically and in terms of application [10]. On the Support Vector Machine, there are several kernels. In this research, the writer uses two kernels as research material, namely linear and polynomial, each of which can be defined by the following functions:

Table 1
Support Vector Machine (SVM)

No	Kernel	Function Definition
1	Linear	$K(x,y) = x \cdot y$
2	Polynomial	$K(x, y) = (x \cdot y + c)^d$

3. Results and Discussion

The sample of tuberculosis (TB) images used in the study consisted of 622 TB images, with 326 negative TB images and 296 positive TB images. Tuberculosis (TB) image samples used as training data amounted to 60% of the 622 images and 40% were used as testing data. The data used in this study are images with a size of 256x256. Image files used for training and testing are taken from [http://openi.nlm.nih.gov/imgs/collections/ChinaSet_AllFiles .zip](http://openi.nlm.nih.gov/imgs/collections/ChinaSet_AllFiles.zip). Examples of some image samples used in this study can be seen in Figure 2 and Figure 3 below.



Fig 2. Sample Negative Tuberculosis (TB)



Fig 3. Sample Positve Tuberculosis (TB)

3.1 Image of HOG Extraction Result

In Figure 4. below, you can see an example of the feature extraction results from HOG:



Fig 4. Example of HOG Extraction

3.2 Testing

In this test, we present the results of the kernel comparison on the SVM using HOG feature extraction in recognizing tuberculosis (TB) x-ray images. The kernels used in this study are linear kernels and polynomial kernels. The total sample used in this study was 600 x-ray images of TB negative and positive TB. The procedure used in the evaluation of TB x-ray image classification in this study is the Apparent Error Rate (APER). The APER value shows the proportion of classification errors by the classification function (Johra, 2018). The explanation of the APER concept can be more clearly identified through the classification error table in Table 1:

Table 2
APER

ACTUAL	PREDICT	
	TRUE	FALSE
TRUE	N_{11}	N_{12}
FALSE	N_{21}	N_{22}

Where the APER value is;

$$APER : = \frac{n_{11}+n_{21}}{n_{11}+n_{12}+n_{21}+n_{22}}$$

The results of the classification of TB x-ray images using the SVM method and Hog feature extraction with polynomial kernels and linear kernels can be seen in table 2 and table 3, with a total sample of 600 TB x-ray images of which 300 TB x-ray images are positive and 300 x -ray negative.



Table 2
Classification of Svm with Polynomial Kernel

ACTUAL	PREDICT		Total Samples X-ray image
	POLYNOMIAL		
	Negative TB	Positive TB	
Negative TB	255	75	300
Positive TB	60	240	300
Accuracy	0.775		

$$\text{APER :} = \frac{75+60}{225+75+60+240} = \frac{135}{660} = 0.225$$

In table 2. It can be seen the classification results on SVM with polynomial kernels in recognizing x-ray images of tuberculosis (TB) using HOG feature extraction, in the first test tested on TB negative x-ray images and the second test on positive TB x-ray images. In testing negative x-ray images, the polynomial kernel TB can recognize 225 x-ray images and 75 errors in recognition, while in positive TB tests can recognize 240 x-ray images and 60 recognition errors. The APER value in the polynomial kernel is 0.225.

Table 3
SVM Classification with Linear Kernel

ACTUAL	PREDICT		Total Sampel Citra X-ray
	LINEAR		
	Negative TB	Positive TB	
Negative TB	201	99	300
Positive TB	24	279	300
Accuracy	0.795		

$$\text{APER :} = \frac{99+24}{201+99+24+279} = \frac{123}{603} = 0.203$$

In table 3. You can see the classification results on SVM with the Linear kernel in recognizing x-ray images of tuberculosis (TB) using HOG feature extraction, in the first test tested on TB negative x-ray images and the second test on positive TB x-ray images. In testing negative x-ray images, the Linear kernel TB can recognize 201 x-ray images and 99 recognition errors, while on positive TB tests can recognize 279 x-ray images and 24 recognition errors. APER value in Linear kernel 0.203

Table 4
Kernel Comparison on SVM in Recognizing Tuberculosis (TB) Disease Images.

KERNEL	Accuracy	APER
POLYNOMIAL	0.775	0.225
LINEAR	0.795	0.203

In table 4, we can see the comparison of the use of the kernel function in SVM in recognizing x-ray images of tuberculosis (TB) using HOG feature extraction, the accuracy obtained in the Polynomial kernel is 0.775 with an APER value of 0.225 while in the Linear kernel it is 0.795 with an APER value of 0.203 . The graph of the classification results of TB images using HOG feature extraction with the SVM kernel function in this study can be seen in Figure 5 below.

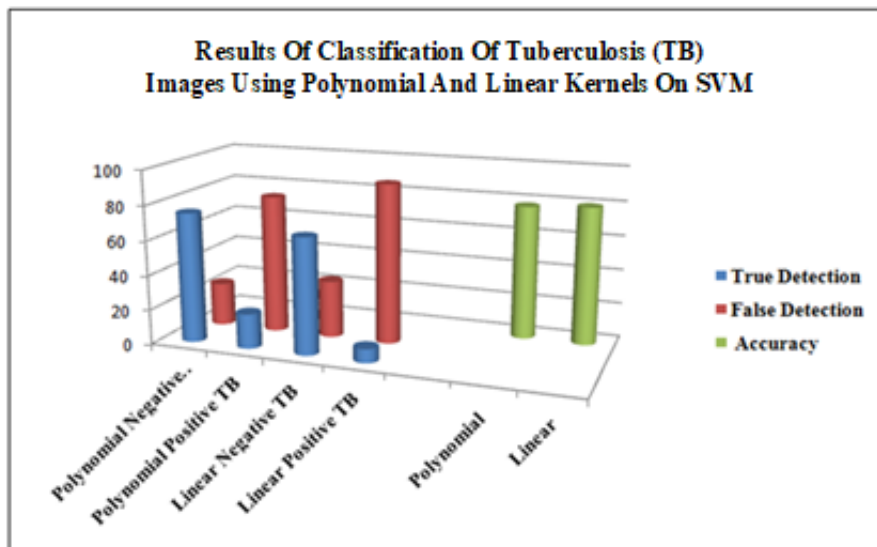


Figure 5. Graph of the results of the classification of tuberculosis (TB) images using the SVM kernel function

4. Conclusion

Based on the results of tests that have been carried out using HOG feature extraction and the SVM kernel function in classifying tuberculosis (TB), the classification results show that the first test was tested on TB negative x-ray images and the second test on positive TB x-ray images. In testing the TB negative x-ray image, the SVM linear kernel accuracy was 67% while the polynomial kernel SVM was 75%, then on the positive TB SVM linear kernel SVM image testing was 92% while the polynomial kernel SVM was 80%. The average accuracy of the polynomial kernel is 77.50% while the linear kernel is 79.50%.

5. References

- [1] Raudlatul Munawarah, Oni Soesanto, and M Reza Faisal, "Penerapan Metode Support Vector Machine Pada Diagnosa Hepatitis," Kumpulan jurnal Ilmu Komputer (KLIK), vol. 4, no. 1, pp. 103-113, 2016.
- [2] Nicole Fogel, "Tuberculosis: A disease without boundaries," Elsevier, pp. 1-5, 2015.
- [3] Moh Yamin Darsyah, "Klasifikasi Tuberkulosis Dengan Pendekatan Metode Supports Vector Machine (Svm)," Statistika, vol. 2, no. 2, pp. 37-41, 2014.
- [4] Muhammad Athoillah, "Pengenalan Wajah Menggunakan SVM Multi Kernel dengan Pembelajaran yang Bertambah," JOIN (Jurnal Online Informatika), vol. 2, no. 2, pp. 84-91, 2017.
- [5] S V Kogilavani, S Malliga, G Jaipriya, Kausalyaa Kausalyaa, and R Makesh Kumar, "Tuberculosis Detection Using Svm And Cnn," International Journal of Advanced Science and Technology, vol. 29, no. 3, pp. 1039-1048, 2020.
- [6] T Karnkawinpong and Y Limpiyakorn, "Classification of pulmonary tuberculosis lesion with convolutional neural networks," International Conference on Computer and Electrical Engineering, pp. 1-5, 2019.
- [7] Putra Apriadi Siregar, Fitriani Pramita Gurning, Eliska Eliska, and Muchti Yuda Pratama, "Analisis Faktor Yang Berhubungan Dengan Kejadian Tuberkulosis Paru Anak Di Rsud Sibuhuan," Jurnal Berkala Epidemiologi, vol. 6, no. 3, pp. 268-275, 2018.
- [8] Yuna Sugianela and Nanik Suciati, "Ekstraksi Fitur Pada Pengenalan Karakter Aksara Jawa Berbasis Histogram Of Oriented Gradient," JUTI, vol. 17, no. 1, pp. 64-72, 2019.
- [9] Ni Luh Made Asri Mulyasari, Nanik Suciati, and Arya Yudhi Wijaya, "Implementasi Deteksi Copy-move Forgery pada," Implementasi Deteksi Copy-move Forgery pada Citra menggunakan Metode Histogram of Oriented Gradients (HOG), vol. 6, no. 2, pp. 449-453, 2017.
- [10] Imelda A Muis and Muhammad Affandes, "Penerapan Metode Support Vector Machine (SVM) Menggunakan Kernel Radial Basis Function (RBF) Pada Klasifikasi Tweet," Jurnal Sains, Teknologi dan Industri, vol. 12, no. 2, pp. 189-197, 2015.
- [11] Sihombing, O., Zendrato, N., Laia, Y., Nababan, M., Sitanggang, D., Purba, W., ... & Siregar, S. (2018). Smart home design for electronic devices monitoring based wireless gateway network using cisco packet tracer. *JPhCS*, 1007(1), 012021.