Indonesian Journal of
Tropical and Infectious Disease
Vol. 3. No. 1 January–March 2012

Literature Review

USING LEARNING VECTOR QUANTIZATION METHOD FOR AUTOMATED IDENTIFICATION OF MYCOBACTERIUM TUBERCULOSIS

Endah Purwanti[1], Prihartini Widiyanti[1,2]
[1] Department of Physics
Science & Technology Faculty Airlangga University, Surabaya, Indonesia
[2] Institute of Tropical Disease
Airlangga University, Surabaya, Indonesia

ABSTRACT

In this paper, we are developing an automated method for the detection of tubercle bacilli in clinical specimens, principally the sputum. This investigation is the first attempt to automatically identify TB bacilli in sputum using image processing and learning vector quantization (LVQ) techniques. The evaluation of the learning vector quantization (LVQ) was carried out on Tuberculosis dataset show that average of accuracy is 91,33%.

Key words: *Mycobacterium Tuberculosis, Image Processing, Learning Vector Quantization.*

INTRODUCTION

In global terms, there are one billion people infected with tuberculosis at any one time. Eight million new cases arise annually making a prevalence of 16 million cases. Three million persons die annually from TB. However, despite these grim figures, even without the influence of treatment and immunisation, the incidence is not as high as it was in the last century. The annual risk of TB infection in South East Asia is 1 to 2.5 per cent. This represents an upward trend. In Indonesia. there are at least half a million new cases of TB per year and 175,000 deaths. Tuberculosis is the second killer of adults after cardiovascular disease and the most important killer out of all the communicable diseases.

At the local level, diagnosis is best achieved through microscopic examination of the bacillus in a sputum. Culturing the bacillus is expensive and impractical as it takes 6 weeks and X-rays can be misleading. Skin testing is recommended by WHO, however, it is not specific for human TB bacillus. Additionally, the size of the reaction is not always helpful as strong reactions may occur in healthy people with repeated occupational exposure to infectious tuberculosis patients and in those with old healed disease.

The purpose of this study is to automatically identify TB bacilli in sputum smears using image processing methods and learning vector quantization (LVQ) techniques. LVQ systems are flexible, easy to implement, and can be applied in multi-class problems in a straightforward fashion. Because LVQ prototypes are determined in the feature space of observed data, the resulting classifiers can be interpreted intuitively. Consequently, LVQ classifiers are widely used in a variety of areas including image processing tasks, medical applications, control of technical processes, or bioinformatics.

THEORY

Tuberculosis

Tuberculosis (TB) is a contagious disease caused by a germ called "Mycobacterium tuberculosis" or "M. tuberculosis". Germs or "bacteria" are tiny living organisms that reproduce by dividing, and can be shaped like a sphere, rod or spiral. They are present virtually everywhere. Some of them are harmless – others are very dangerous.

TB is manly spread by airbone transmission. The source of infection is a patient with pulmonary (or laryngeal) TB who expectorates bacilli. During coughing, speaking, or sneezing, the patient produces tiny infectious droplets; these droplets dry out and remain in the air for several hours. Contamination occurs when these infectious droplets are inhaled. Sunlight and ventilation are effective in decontaminating the environment.

The infectiousness of a patient is linked to quantity of bacilli contained in his/her sputa. In adults and older children: sputum obtained spontaneously. Sputum emitted in early morning often shows a higher concentration of *M. tuberculosis*. Sputum smear microscopy allow a simple, rapid and reliable identification of patients, but has a low sensitivity. The reliability of this examination depends also on the proper preparation and interpretation of slides. Quality control checks must be regularly carried out in the laboratory.

Image Processing

The image processing techniques used in this investigation help in enhancing the images and highlighting features needed for the shape description of each bacillus in the image. Separate entities in the images, such as TB bacilli, will be referred to as objects (each object being considered as a separate region).

1. *Image Enhancement*. The main goal of image enhancement is to process an image in some way so as to render it more visually acceptable or pleasing. Some techniques of image enhancement are removal of noise, sharpening of image edges and 'soft focus' (blurring) effect.
2. *Edge Detection.* Edges are simply regions of intensity transition between one object and another.One of edge detection method is Canny edge detector. Canny aimed to develop an edge detector that satisfied three key criteria: A low error rate. In other words, it is important that edges occuring in images should not be missed and that there should be no response where edges do not exist. The detected edge points should be well localized. In other words, the distance between the edge pixels as found by the detector and the actual edge should be a minimum. There should be only one response to a single edge.
3. *Morfhological processing*. The aim is to identify and extract meaningful image descriptors based on properties of form or shape within the image. Morphological operations can be applied to images of all types, but the primary use for morphology (or, at least, the context in which most people will first use it) is for processing binary images and the key morphological operators are the relatively simple ones called dilation and erosion.
4. *Feature Extraction*. The aim is to process the image in such a way that the image, or properties of it, can be adequately represented and extracted in a compact form amenable to subsequent recognition and classification.

Learning Vector Quantization

The LVQ classifier is based on the principle of nearest neighbor, which is demonstrated in the Figure. 1, where Euclidean distance is basically used for calculating distance.

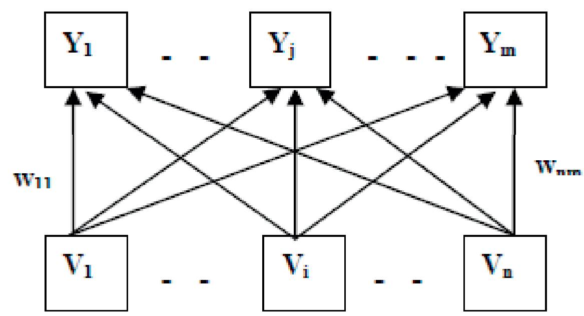


Figure 1. LVQ Architecture

Consider a sequences of labeled input samples {Xi: Xi Є R^n, i = 1, 2,…,N}; each sample input xi is tagged with its "correc class. Let t denote the number of training iterations. Assume that a set of cluster vectors {Ci: Ci Є R^n, i = 1, 2,…,m; m<<N}; have been obtained. The computing similarity and selecting criteria are unmodified but the adaptation is given by firs finding two closest cluster Ci and Cj vector to X and assume their distances to X are S1 and S2 respectively. Correction are made, two cases:

Case 1. Ci and Cj belong to different Classes, but one of them is correct.

If (S1/S2) > [(1-W)/(1+W)]
Ci = Ci-ά(t)[X-Ci]
Cj = Cj+ ά(t)[X-Cj]

Where X and Cj belong to the same class, while X and Ci belong to different classes.

Case II. Ci and Cj belong to same classes

Ck = Ck + e ά(t)[X-Ck]
For k Є {i,j} and X, Ci, Cj belong to the same class.

METHODOLOGY

In the proposed method, we first pre-processed the image of sputum using median filter. The median filter is used to preserving sharp high-frequency detail (i.e. edges) whilst also eliminating noise with minimal degradation or loss of detail in the image. Second, we partitioned the image into objects of defective areas, by performing Watershed Segmentation method. In watershed segmentation, we envisage the 2-D, grey-scale image as a topological surface or 'landscape' in which the location is given by the x,y image coordinates and the height at that location corresponds to the image intensity or grey-scale value. Third, we extracted the image properties by using single parameter shape descriptors. The aim is simply to characterize a shape as succinctly as possible in order that it can be differentiated from other shapes and classified accordingly. These features extracted are fed to the LVQ, which is a supervised learning method of kohenon network, which uses competitive learning technique. The objective of LVQ based training and classification method performed, is to develop software to improve the process of automated identification of mycobacterium tuberculosis. The proposed system is having five different phases, which is shown in the following Figure 2.

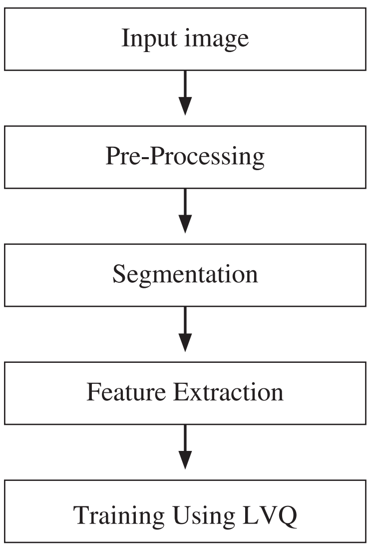


Figure 2. Design System

Training algorithm for LVQ
Steps:

1. Set the initial value of weight matrix and learning rate.
2. Execute steps from 3 to7 until end condition is true.
3. Execute steps from 4 to 5 for each 'V', where 'V' is the input vector.
4. Using squared Euclidian Distance calculate J as follows
   a. D(j) = ý(*wij* Σ *Vi*), where D(j) is squared Euclidian Distance wij is the element of weight matrix Vi is the element of input vector
   b. Find J with D(j) is minimum
5. Wj is updated as follows:
   a. if (T = Cj) then
      T is the target and Cj is winner index
      – wj(N) = wj(O) + α (V-wj(O))
   b. if (T! = Cj) then
      – wj(N) = wj(O) - α (V-wj(O))
   where,wj(O) is old value of wj wj(N) is new value of wj
6. Decrement the learning rate
7. Check for end condition, which may be fixed number of iterations.

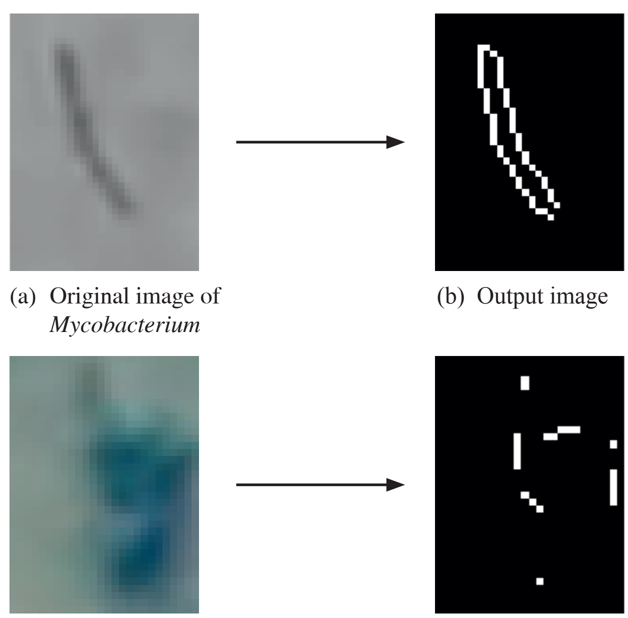
The weights are updated on each step for the process of learning. The weights will move closer to the class which will be winning class or else it will move away from the class. Once the training is finished, the LVQ will be able to recognize any unknown features which are not trained already.

RESULT

We used the 60 images of sputum in learning vector quantization system. The data was divided into a training set and a test set. A training set is 20 image of sputum which consist of 10 images *Mycobacterium* and 10 images are not. A test set of 40 objects (bacilli and non-bacilli) were used to test the system.

The input images are first passed through preprocessing stages. Watershed method is used for segmentation which gave a better result than region growing. The output of segmentation process are shown in the Figure. 3. Feature extraction process is a very important step where in proposed method for our application.

In the training phase, the learning vector quantization was applied on training data. Then, for data in the testing set, classification process seaches in this LVQ for finding the class that is closest to be attached with the object presented for categorization. Figure 4 present the experimental result for LVQ classifier. The figure show that the LVQ classification was successful for identification of Mycobacterium tuberculosis. The evaluation of the learning vector quantization (LVQ) was carried out on Tuberculosis dataset show that average accuracy is 91,33%.



(a) Original image of *Mycobacterium* (b) Output image

(c) Original image of *Not Mycobacterium* (d) Output image

Figure 3. The sample output of segmentation process

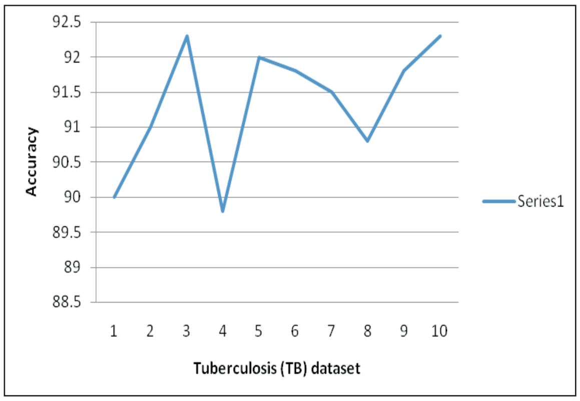


Figure 4. Accuracy LVQ classifier

CONCLUSIONS

In this paper, we presented learning vector quantization method applied to automatically identify TB bacilli. The evaluation of the learning vector quantization (LVQ) was carried out on Tuberculosis dataset show that average of accuracy is 91,33%.

REFERENCES

1. Bakar AA, Febriyani F, 2007. Rough Neural Network Model for Tuberculosis Patient Categorization. Proceedings of the International Conference on Electrical Engineering and Informatics Institut Teknologi Bandung, Indonesia.
2. Deserno TM, 2010. Biomedical Image Processing, Springer.
3. Dougherty G., 2009. Digital Image Processing for Medical Application, Cambridge University Press.
4. Fausett L., 1994. Fundamental of Neural Network. Prentice Hall Inc.
5. Gonzales R, Wintz P., 1987. Digital Image Processing, Second Edition. Addison-Wesley Publishing Company.
6. Kohonen T, 1995. "LVQ_PAK Version 3.1 The Learning Vector Quantization Program Package", Prepared by the LVQ Programming Team of the Helsinki University of Technology.
7. Rao VB, Rao HV, 1993. C++ Neural Networks and Fuzzy Logic, MIS: Press.
8. Siang JJ, 2005. Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan Matlab, Andi Yogyakarta.
9. Solomon C & Breckon T, 2011. Fundamental of Digital Image Processing, Wiley Online Library.
10. Subagyo A. Aditama YT. *et. al.* 2006. Pemeriksaan Interferongamma Dalam Darah Untuk Deteksi Infeksi Tuberkulosis. Bagian Pulmonologi dan Ilmu Kedokteran Respirasi FKUI – RS Persahabatan, Laboratorium Patologi Kinik RS Persahabatan, Jakarta. Jurnal Tuberkulosis Indonesia Vol. 3 No. 2.
11. Thorn P, 2007. Overcoming Tuberculosis: Handbook for Patients, London, England.
12. Tuberculosis General Information, 2007. www.tbc.gov/tb
13. Vaithiyanathan, Venkataraman, Raj A. 2011. Weld Defect Detection in Radiograpy Based on Projection Profile and Invariant by Using LVQ, JATIT & LLS.
14. Veropoulus K, Campbell C. 1998. The Automated Identification of Tubercle Bacilli using Image Processing and Neural Computing Techniques, Proceedings of the 8th International Conference on Artificial Neural Network vol 2.
15. Zain MY, and Nasution TMA, 2009. Identifikasi Bakteri Tuberculosis Berdasar Ciri Morfologi Dan Warna. ITS-Undergraduate 10581-Paper.pdf.