

## Classification of Medical Images Using Support Vector Machine

Vanitha.L.<sup>1+</sup> and Venmathi.A.R.<sup>2+</sup>

<sup>1</sup>Department of Electronics & Communication Engineering, Sri Venkateswara College of Engineering, Sriperumbadur, Chennai, India.

<sup>2</sup>Department of Electronics & Communication Engineering, Kings Engineering College, Irungattukottai, Chennai, India.

**Abstract.** The proposed work is to develop an automatic tool to identify microbiological types without human supervision. Generally Bacteriophage (phage) typing & Fluorescent imaging methods are used to extract representative feature profiles and bacterial types are identified by human experts by reading the feature profiles. These methods are time consuming and prone to errors. In this proposed work, the features of the bacterial image are extracted and Support Vector Machine (SVM) is used for classifying the Bacterial types. Support Vector Machines have high approximation capability and much faster convergence.

**Keywords:** Bacteria, Support Vector Machine, feature extraction.

### 1.Introduction

Bacteria are considered to be the main cause for any disease. For an effective and preventive treatment it needs a characterization of the disease outbreak by identifying its pathogens. The identification of pathogens in a bacteria level or bacteria-species level is not satisfactory for epidemiological and clinical concerns [4], in particular due to increasing bacterial adaptation to human environments including resistance of bacteria to antimicrobial agents. Therefore, a bacterial type diagnosis is required, i.e., the identification of pathogens below the species level [1]. This yields information for controlling the disease.

Sub groupings of bacterial species to types (bacterial types) are used for many important pathogenic bacteria [2] such as the Tuberculosis bacilli, the Staphylococcus aureus (S. aureus), Salmonellae, Vibriones, E-Coli etc. To date, many of the biological and manual procedures are prone to large variability within and across the human experts due to natural fuzziness present in the microbiological data. These procedures are time consuming and are of great cost.



ULCER BACTERIA

YELLOW  
TUBERCULOSIS

RED ALTRUISTIC  
BACTERIA

SKIN BACTERIA

<sup>+</sup> Corresponding author. Tel.:9444752415  
E-mail address: -vvanetha@yahoo.com,vanitha@svce.ac.in  
arvenmathi@yahoo.co.in



**PLAGUE BACTERIA**

**GUT BACTERIA**

Fig. 1 Types of Bacteria

Reducing the amount of human intervention in the data analysis is crucial in order to cope with the increasing volume of data and to achieve more objective and quantitatively accurate measurements as well as to obtain repeatable results [3]. In view of the above, in this work, the image analysis is combined with pattern recognition tools in a general framework and it is focused on microbiological data and bacterial type identification.

The bacterial type identification will be accomplished using Image processing and Pattern matching techniques [5]-[9]. In this work, the SVM one of the most competitive techniques of the popular artificial neural networks is used for validating and testing the bacterial images and for classification.

## 2. Methodology

Figure 2 shows the block diagram of Pattern Recognition System. The main aim of pattern recognition is the classification of patterns and sub patterns in an image. A pattern recognition system includes:

- Subsystems to define pattern class
- Subsystem to extract selected features
- Subsystem for classification known as classifier.

The classifier used in the proposed system is the Support Vector Machine (SVM)



Figure 2. Block diagram of Feature extraction System

### 2.1. Modules

In the proposed system there are three modules:

- Pre-processing
- Feature extraction
- Bacteria Classification using SVM classifier

### 2.2. Pre-Processing

The purpose of pre-processing is to remove the noise from the image. This is required for the reliable extraction of features as feature extraction algorithms give poor results in the presence of a noisy background.

### 2.3. Feature Extraction

The three main approaches for feature extraction and classification based on the type of features are as follows:

- statistical approach
- syntactic or structural approach
- Spectral approach.

In case of statistical approach, pattern is defined by a set of statistically extracted features represented as vector in multidimensional feature space. In case of syntactic approach, texture is defined by texture primitives, which are spatially organized according to placement rules to generate complete pattern. In syntactic pattern recognition, a formal analogy is drawn between the structural pattern and the syntax of language.

In case of spectral method, textures are defined by spatial frequencies and are evaluated by autocorrelation function of a texture.

As a comparison between the above-mentioned three approaches, spectral frequency-based methods are less efficient, while statistical methods are particularly useful for random patterns; while for complex patterns, structural methods give better results.

The features that are extracted from Bacterial image are Relative length, Relative area, Mean, Standard deviation, Entropy, Eccentricity and length to width ratio.

## 2.4. Bacterial images classifier

Bacterial classification is performed using Support Vector Machine as a classifier.

## 3. Support Vector Machine (SVM)

SVM has proven its efficiency over neural networks and RBF classifiers. Unlike neural networks, this model builds does not need hypothesizing number of neurons in the middle layer or defining the centre of Gaussian functions in RBF [11]. SVM uses an optimum linear separating hyperplane to separate two set of data in a feature space. This optimum hyperplane is produced by maximizing minimum margin between the two sets [12]. Therefore the resulting hyperplane will only be depended on border training patterns called support vectors.

The support vector machine operates on two mathematical operations: (1) Nonlinear mapping of an input vector into a high-dimensional feature space that is hidden from both the input and output. (2) Construction of an optimal hyperplane for separating the features discovered in step 1. Support vectors are determined by using the equations 1 -8. [10, 11]

### 3.1. Variable definition:

1. Let  $x$  denote a vector drawn from the input space, assumed to be of dimension  $m_0$ .
2. Let  $\{\phi_j(x)\}$  for  $j=1$  to  $m_1$ , denote a set of nonlinear transformations from the input space to the feature space.
3.  $m_1$  is the dimension of the feature space.
4.  $\{w_j\}$  for  $j=1$  to  $m_1$  denotes a set of linear weights connecting the feature space to the output space.
5.  $\{\phi_j(x)\}$  represent the input supplied to the weight  $w_j$  via the feature space.
6.  $b$  is the bias
7.  $\alpha_i$  is the Lagrange coefficient
8.  $d_i$  corresponding target output

### 3.2. Steps involved in the design of SVM:

1. Hyperplane acting as the decision surface is defined as

$$\sum_{i=1}^N \alpha_i d_i K(x, x_i) = 0 \quad (1)$$

Where

$K(x, x_i) = \phi^T(x)\phi(x_i)$  represents the inner product of two vectors induced in the feature space by the input vector  $x$  and input pattern  $x_i$  pertaining to the  $i$ th example. This term is referred to as inner-product kernel.

where

$$w = \sum_{i=1}^N \alpha_i d_i \phi(x_i) \quad (2)$$

$$\phi(x) = [\phi_0(x), \phi_1(x), \dots, \phi_{m_1}(x)]^T \quad (3)$$

$\varphi_0(x) = 1$  for all  $x$

$w_0$  denotes the bias  $b$

2. The requirement of the kernel  $K(x, x_i)$  is to satisfy Mercer's theorem. The kernel function is selected as a polynomial learning machine.

$$K(x, x_i) = (1 + x^T x_i)^2 \quad (4)$$

3. The Lagrange multipliers  $\{\alpha_i\}$  for  $i = 1$  to  $N$  that maximize the objective function  $Q(\alpha)$ , denoted by  $\alpha_{0,i}$  is determined.

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(x_i, x_j) \quad (5)$$

Subject to the following constraints:

$$\sum_{i=1}^N \alpha_i d_i = 0 \quad (6)$$

$$0 \leq \alpha_i \leq C \quad \text{for } i = 1, 2, \dots, N \quad (7)$$

4. The linear weight vector  $w_0$  corresponding to the optimum values of the Lagrange multipliers are determined using the following formula:

$$w_0 = \sum_{i=1}^N \alpha_{0,i} d_i \varphi(x_i) \quad (8)$$

$\varphi(x_i)$  is the image induced in the feature space due to  $x_i$ .

$w_0$  represents the optimum bias  $b_0$ .

## 4. Implementation and Results

### 4.1 Input Data:

The input to the feature extraction algorithm is the bacterial images. The pattern vectors (features) extracted from the images is given as input to the SVM classifier. Large database are required for the classifier to perform the classification correctly. In this system a sample of 180 bacterial images are collected and 120 images are used for training phase and remaining 60 images are used for testing. The Classification accuracy and error rate is obtained by using the following formula:

Number of correctly classified Bacterial images

$$\text{Accuracy} = \frac{\text{Number of correctly classified Bacterial images}}{\text{Total number of bacterial images}} \quad \text{-- (9)}$$

Number of misclassified Bacterial images

$$\text{Error Rate} = \frac{\text{Number of misclassified Bacterial images}}{\text{Total Number of bacterial images}} \quad \text{-- (10)}$$

### 4.2 Result:

The following table shows the number of input images for training phase and testing phase and the correctly classified images by Support Vector Machine.

Type	No. of input images	Correctly classified Images	Efficiency (%)

	Training Phase	Testing Phase	Training Phase	Testing Phase	Training Phase	Testing Phase
1	20	10	20	10	100	100
2	20	10	19	9	95	90
3	20	10	20	9	100	90
4	20	10	20	10	100	100
5	20	10	18	8	90	80
6	20	10	20	10	100	100
Total	120	60	117	56	97.5	93.33

Table 1 Number of correctly classified bacteria images by SVM.

## 5. Conclusion

The proposed approach using SVM as a classifier for classification of Bacterial images provides a good classification efficiency of 97.5% during training phase and 93.33% efficiency during testing phase. The proposed approach is computationally less expensive and yields good result. The classification accuracy can be improved by extracting more features and increasing the training data set.

## 6. References

- [1] Guisong Liu and Xiaobin Wang, "An Integrated Intrusion Detection System by using Multiple Neural Networks", IEEE, pp22-28, 2008
- [2] Syed Zahid Hassan and Brijesh Verma, "A Hybrid Data Mining Approach for Knowledge Extraction and Classification in Medical Databases", Seventh International Conference on Intelligent Systems Design and Applications, pp.503-507, 2007
- [3] T.G. Emori and R.P. Gaynes, "An overview of nosocomial infections, concluding the role of the microbiology laboratory," Clin. Microbiol. Rev., Vol. 6, no. 4, pp. 428–442, 1993.
- [4] S.S. Purohit, "Microbiology Fundamentals Applications" Publisher, India, sixth edition, 2003.
- [5] G. Teper, G. Ziv, and E. Skutelski, "Flowcytometry analysis of S.aureus– Bacteriophage interactions," in Proc 3rd Int. Mastitis Sem., vol, As-1, p-8, 1995.
- [6] T.G. Emori and R.P. Gaynes, "An overview of nosocomial infections, concluding the role of the microbiology laboratory," Clin. Microbiol. Rev., Vol. 6, no. 4, pp. 428–442, 1993.
- [7] S.S. Purohit, "Microbiology Fundamentals Applications" Publisher, India, sixth edition, 2003.
- [8] G. Teper, G. Ziv, and E. Skutelski, "Flowcytometry analysis of S.aureus– Bacteriophage interactions," in Proc 3rd Int. Mastitis Sem., vol, As-1, p-8, 1995.
- [9] B.D. Singh, "Genetics", Kalyani Publishers, 2005. Earl Gose, Richard Johnson-baugh, Steve Jost, "Pattern recognition and image analysis", Prentice Hall of India Private Limited, New Delhi, 2000
- [10] Laurene Fausett, Fundamentals of Neural Networks, Pearson Education, 2007
- [11] G. Angiulli, M. Cacciola, and M. Versaci, Microwave Devices and Antennas Modelling by Support Vector Regression Machines, IEEE Transactions on Magnetics, Vol 43, April 2007
- [12] Nurhan Turker Tokan, Filiz Gunes, Analysis and Synthesis of the Microstrip Lines Based on Support Vector Regression, Proceedings of the 38th European Microwave Conference