BAYESIAN PROKARYOTE CLASSIFICATION FROM MICROSCOPIC IMAGES

Amaleena Mohamad, Noorain A. Jusoh, Zaw Zaw Htike and Shoon Lei Win

Faculty of Engineering, IIUM, Kuala Lumpur, Malaysia

ABSTRACT

Great knowledge and experience on microbiology are required for accurate bacteria identification. Automation of bacteria identification is required because there might be a shortage of skilled microbiologists and clinicians at a time of great need. We propose an automatic bacteria identification framework that can classify three famous classes of bacteria namely Cocci, Bacilli and Vibrio from microscopic morphology using the Naïve Bayes classifier. The proposed bacteria identification framework comprises two steps. In the first step, the system is trained using a set of microscopic images containing Cocci, Bacilli, and Vibrio. The input images are normalized to emphasize the diameter and shape features. Edge-based descriptors are then extracted from the input images. In the second step, we use the Naïve Bayes classifier to perform probabilistic inference based on the input descriptors. 64 images for each class of bacteria were used as the training set and 222 images consisting of the three classes of bacteria and other random images such as humans and airplanes were used as the test set. There are no images overlapped between the training set and the test set. The system was found to be able to accurately discriminate the three classes of bacteria. Moreover, the system was also found to be able to reject images that did not belong to any of the three classes of bacteria. The preliminary results demonstrate how a simple machine learning classifier with a set of simple image-based features can result in high classification accuracy. The preliminary results also demonstrate the efficacy and efficiency of our two-step automatic bacteria identification approach and motivate us to extend this framework to identify a variety of other types of bacteria.

KEYWORDS

Bacteria Identification, Cocci, Bacilli, Vibrio, Naïve Bayes, Machine Learning

1. INTRODUCTION

Bacteria, which are prokaryotic microorganisms, are the most abundant and simplest organisms in the world as we know it. Prokaryotes do not possess a nucleus and complex organelles. Because most prokaryotes range in size less than ten micrometers (µm), microscopes are used to study bacteria. Bacteria identification is very important in microbiology and pathology as it serves a basis of understanding diseases. Due to this, various types of methods have been introduced to classify bacteria in microbiology. Clinicians and microbiologists commonly employ the typing schemes which are dependent on the phenotypic typing schemes to develop the bacterial morphology and staining properties of the organism. Vectors, environmental reservoir of organism and pathogen's ways of transmission is important for the clinicians. Therefore, it is extremely essential to perform bacteria classification such that the said information can be obtained. On the other hand, scientists who are interested in microorganisms' evolution are getting more interested in taxonomic techniques which permit the comparison of highly

conserved genes among dissimilar species. Therefore, computerized techniques are required for this task [6-18].

The most basic technique used for classifying bacteria is based on the bacterium's shape and cell arrangement. The most ordinary shapes of bacteria include rod, cocci (round), and spiral forms. Cellular arrangements occur singularly, in series, and in groups. Some species have one to numerous projections called flagella which enable the bacteria to swim and move. Cocci or coccus for a single cell are round cells, occasionally flattened when being adjacent to each other. Cocci bacteria can exist individually, in pairs, in groups of four, in chains, in clusters or in cubes consisting of eight cells. Bacilli are rod-shaped bacteria which also can occur individually, in pairs, or in chains. Figure 1 shows examples of three famous classes of bacteria namely: cocci, vibrio and bacilli.
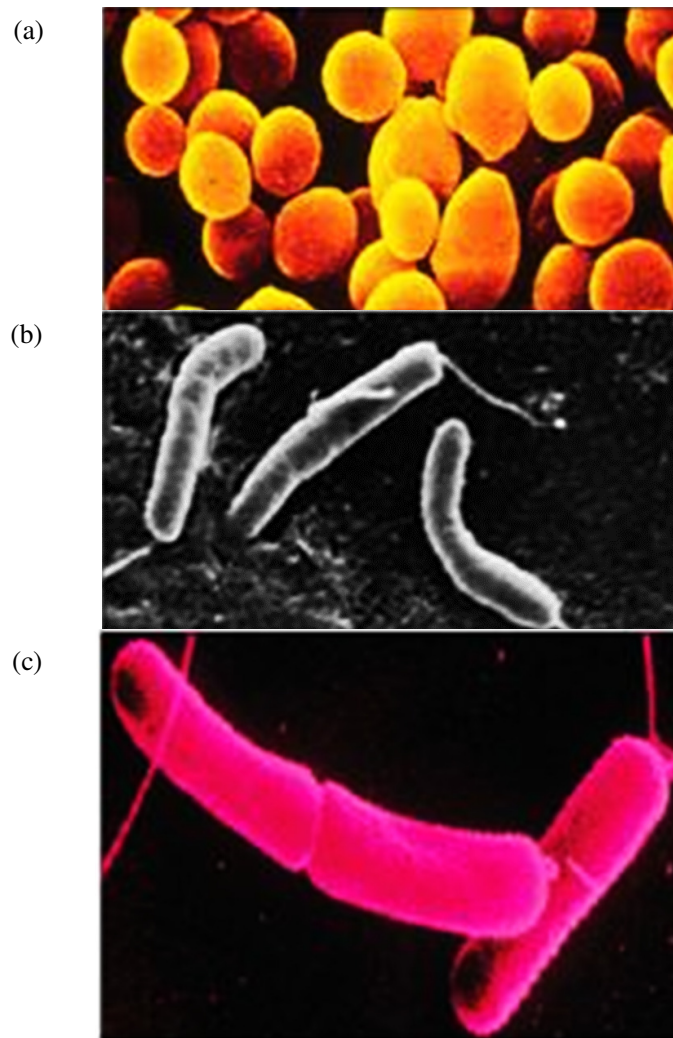

Figure 1. Three classes of bacteria (a) Cocci (b) Vibrio (c) Bacilli [21]

2. RELATED WORK

Bacteria classification plays important role in yielding information for disease control. Bacterial species are usually sub-grouped to different types and is used for many crucial pathogenic bacteria such as Salmonellae, E Coli, and Vibriones [1]. H.C. Gram in 1884 discovered the Gram stain classification remains an important and useful technique until today. This technique classifies bacteria as either Gram positive or negative based on their morphology and differential staining properties [2]. Table I shows the general phenotypic classification of bacteria for the Gram positive [3].

Other types of bacteria classifications that are commonly used are based on the prokaryotes of the bacteria which include their function and structures such as the slime, capsule, peptidoglycan, cytoplasmic membrane, flagella, pili, and the secreted products. The main objectives of phenotypic classification are to generate clusters of strains and establish the hierarchy of bacteria species as members of different species may share high levels of resemblance [3].

Increasing bacterial adaptation level to human environments proves that the identification of pathogens in a bacteria species level is unsatisfactory [1].Therefore bacterial type diagnosis is required, which involves the classification of pathogens below species level. To date, biological and microbiological data analysis entails an extensive amount of human intervention [1]. The manual procedures are susceptible to inconsistency and are a tedious and complicated work which needs abundant correlative data. [3].These procedures also consume a lot of time and energy and are of great cost [1]. It is therefore important to reduce the amount of human intervention in order to handle the rising data volume besides trying to achieve adequate level data accuracy. Due to inefficient manual procedures in classifying bacteria, a considerable amount of research has been done on automatic bacteria classification using various approaches which include pattern recognition technique.

Table 1: General Phenotypic Classification of Bacteria [1]

Name	Morphology	Type of infection
Staphylococci	Cocci in grapelike clusters	Soft tissue, bone, joint, food poisoning
Streptococci	Cocci in pairs, chains	Skin pharyngitis, endocarditis, toxic, shock
Enterococci	Cocci in pairs, chains	UTI, GI, catheterrelated, infections
Bacilli	Rods, sporeforming	Anthrax, food poisoning, catheter-related infections
Clostridia	Rods, spore formers	Tetannus, diarrhea, gas gangrene

Pattern recognition which is a field of machine learning is an act of taking in raw data and taking action based on pattern category [4-5]. The algorithm of pattern recognition normally consists of image segmentation, feature extraction, and classification [5]. Desired object is first segmented and the selected features are extracted from the image. Then, the classifier can be trained using these features as input data. Example of algorithm which can be used in bacteria classification is shown in Figure 1[5].
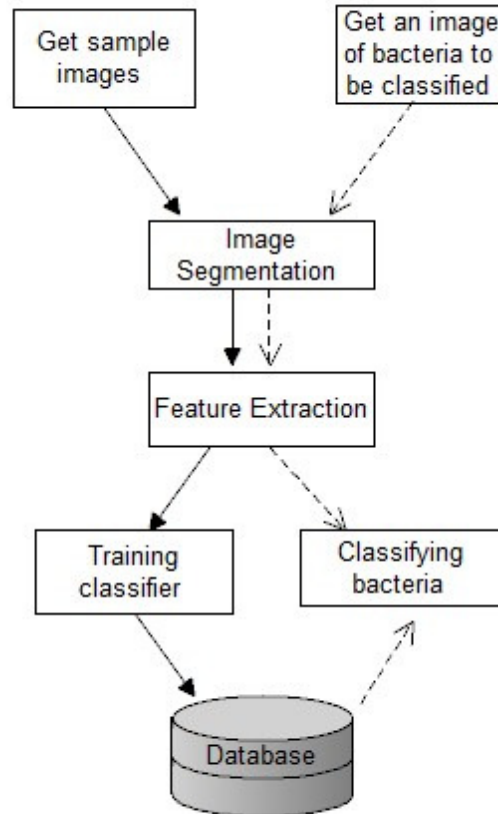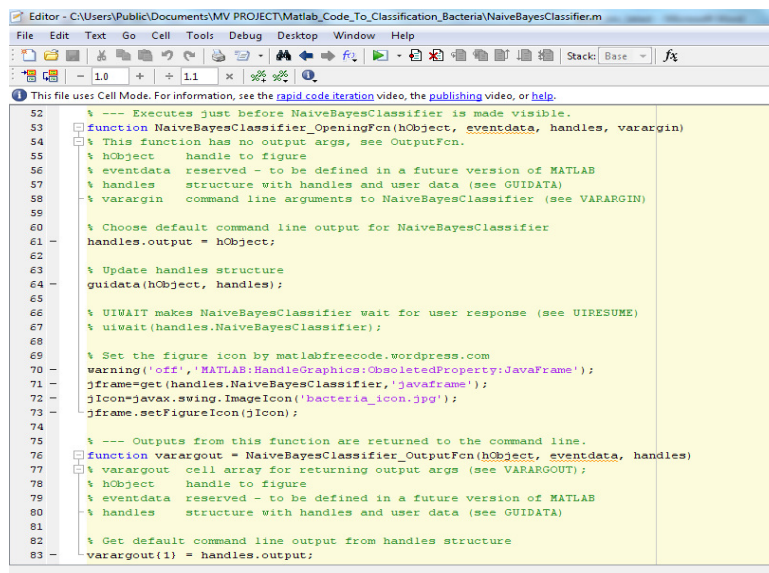
Figure 2. Algorithm used in bacteria classification technique.

Traditionally, bacteria classification is carried out manually, making the task to be extremely tedious, complicated, and susceptible to inconsistency and inaccuracy. It also consumes a lot of time and requires constant presence of specialists so as to avoid mistakes during classifying. Accurate and fast bacteria classification is extremely important as it may help to perform early diagnosis of severe diseases thus making it possible to be controlled. Therefore, it is very important to eliminate this manual procedure and introduce a new technique on bacteria classification using pattern recognition in order to achieve more accurate result and to make it faster, and more efficient. Existing techniques in the literature have do not produce satisfactory results in bacteria identification. The Naïve Bayes classifier in machine learning has well-known resilience to irrelevant and a high number of features. In the task of bacteria identification from microscopic morphology, there can be a number of irrelevant features since input images are highly dimensional. Therefore, we propose the Naïve Bayes classifier to solve this problem.

3. PROPOSED FRAMEWORK

We propose edge detection in conjunction with a naïve Bayes classifier to perform bacteria classification. As with any machine learning problem, our proposed framework comprises two steps: feature extraction and classification. For feature extraction, based on the domain knowledge of some microbiologists, two principal features of bacteria, which are shape and diameter are selected. In order to extract features, Canny edge detection technique is first employed. Next, shape-based descriptors are extracted. The next step is to train and classify our bacteria images accordingly. For this purpose, Naive Bayes Classifier is applied. Naïve Bayes classifiers is chosen to be applied in this system as it only needs a little number of training data to approximate the factors required for classification process. Furthermore, it is easy and fast to build the model despite being able to be modified with new set of training data without the needs to rebuild the model. In another words, it is advantageous and simple to be applied. 64 images for each class of bacteria, cocci, bacilli and vibrio are chosen as training data set to train the proposed bacteria identification system. The proposed framework was implemented on Matlab as shown in Figure 5.

Figure 5. Initialization of Naive Bayes Classifier in Matlab

During training process, the training images are loaded and trained using the GUI. To test our classifier, we create a validation set comprising 222 images consisting of the three classes of bacteria and other random images such as humans and airplanes. Figure 6 shows the sequence of procedures to train and test the bacteria classification system. The steps are as follows:

1. Naive Bayes Classifier interface is executed
2. Training data are loaded and trained based on their shape and diameter measurement
3. Testing data are loaded and tested
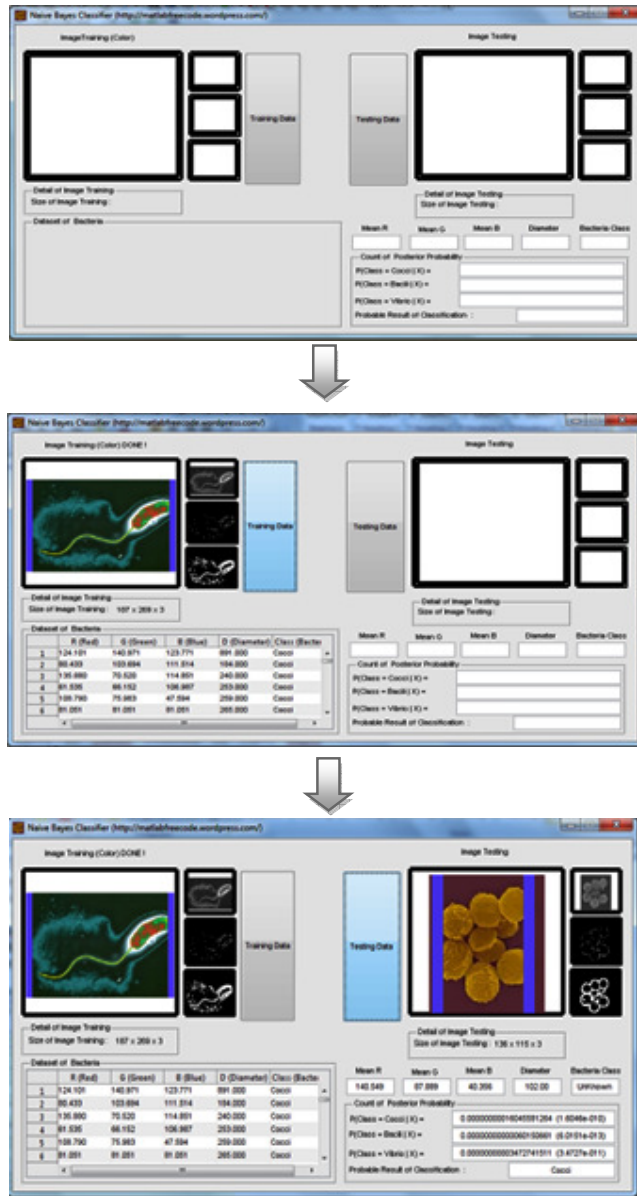4. Classification results are displayed

Figure 6. GUI to train and test the bacteria classification system

4. EXPERIMENTS

Using the training set containing 64 images and test set containing 222 images, we obtained an accuracy of 81.08%. Results show that the easiest bacteria image to be classified is the Bacilli followed by Vibrio and finally the Cocci. Vibrio and Bacilli are successfully classified using this system; however for the cocci, about forty percent of the total images show unknown results after being tested by the system. Figures 7, 8, and 9 show the results for the classification of the three bacteria classes. Negative results are obtained when system is tested with unrelated image as shown in figure 10. Edge detection for segmentation helps a lot in improving the accuracy of the bacteria classification system besides applying the naïve Bayes Classifier. Probabilistic results are

also displayed in the classifier interface as to show the relation between system performance and its accuracy. The results demonstrate how a simple naïve Bayes Classifier with just a few edge-based features can achieve high classification accuracy.
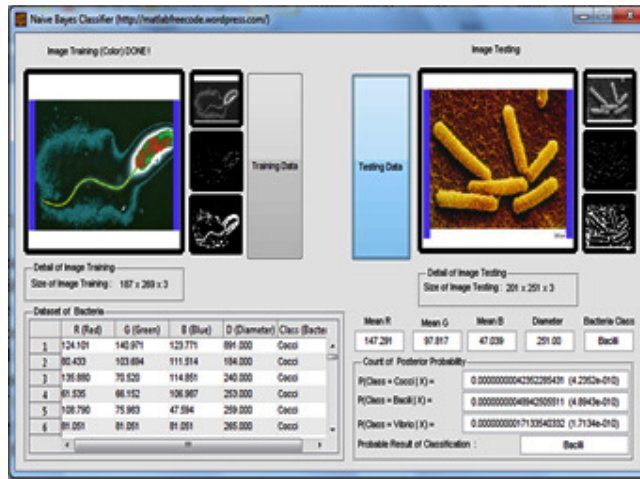

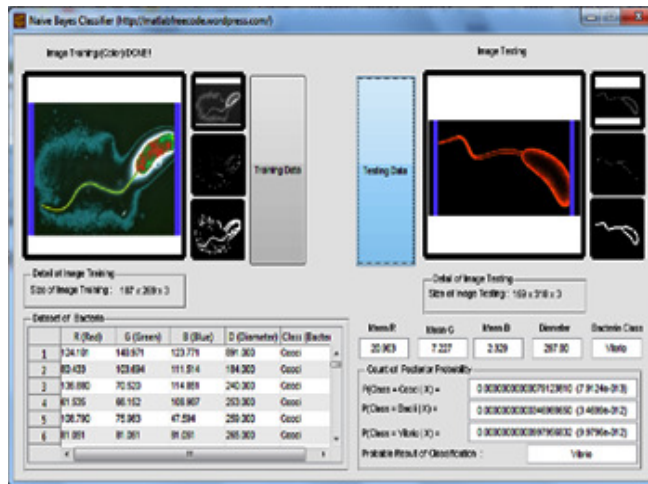Figure 7. Positive result when tested using Bacilli Image

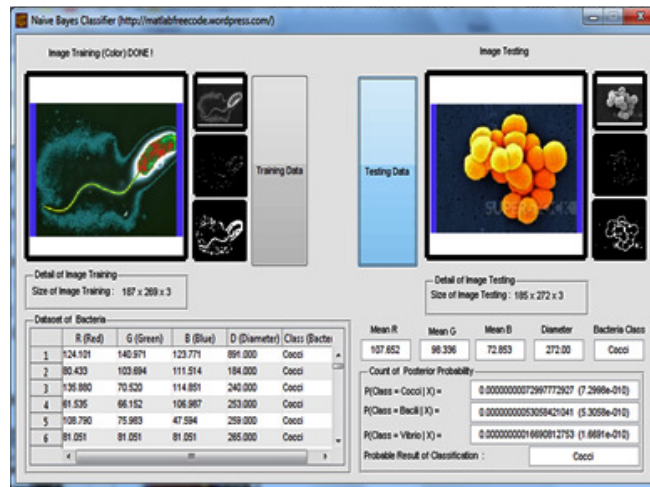Figure 8. Positive result when tested using Vibrio Image

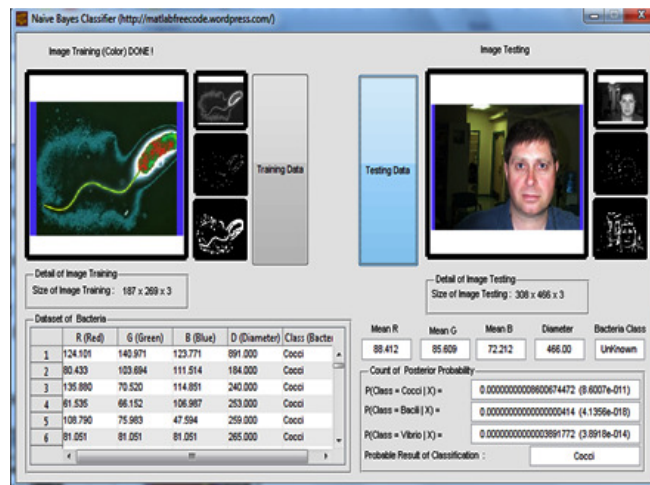Figure 9. Positive result when tested using Cocci Image

Figure 10. Negative result when testing using Unrelated Image

5. CONCLUSION

This paper demonstrates how a simple machine learning classifier with a few simple image-based features can render high classification accuracy in bacteria identification from microscopic morphology. A two-step bacteria identification framework, which consists of feature extraction and classification, is proposed. The proposed system has achieved over 80% accuracy in classifying three major classes of bacteria namely: Cocci, Bacilli, and Vibrio. Despite its exploratory nature, this study has achieved high classification accuracy and offers some insight into this machine learning area. Considerably more work will need to be directed in this area to achieve highly robust, and accurate classification using machine learning, not only for bacteria, but for many other object classification. As future work, we would like to extend this framework to classify a variety of different types of bacteria.

REFERENCES

[1] S.Prabakar, K.Porkumaran, J. Samson Isaac, "Development of Image Processing Scheme for Bacterial Classification Based On Optima Dicriminant Feature" presented at 2010 International Conference of Bioinformatics and Biometrics Engineering.

[2] L.Frank, "Bacterial Classification, Structure and Function", 2009, pp.1-2

[3] Z. Ying, W. Zhiye and Z. Jianping., "Bacteria Classification Using Neural Network", Institute of Biology Gansu Academy of Sciences, 2011.

[4] O.Richard , E. Peter and G. David, "Pattern Classification", 2nd ed, Wiley-Interscience, 2000

[5] Y. Huiyong, L.Wei, X. Kun,Q.   Jian and W.Xin, "Research on Insect Identification   Based on Pattern Recognition Technology", Taiyuan Extension Station of Application Technology, 2010

[6] Z.Z. Htike, S.L. Win "Recognition of Promoters in DNA Sequences Using Weightily Averaged One-dependence Estimators", Procedia Computer Science, Volume 23, 2013, Pages 60-67, ISSN 1877-0509.

[7] Z.Z. Htike, S.L. Win "Classification of Eukaryotic Splice-junction Genetic Sequences Using Averaged One-dependence Estimators with Subsumption Resolution", Procedia Computer Science, Volume 23, 2013, Pages 36-43, ISSN 1877-0509.

[8] Z.Z. Htike, S. Egerton, Y.C. Kuang, "A Monocular View-Invariant Fall Detection System for the Elderly in Assisted Home Environments," 7th International Conference on Intelligent Environments (IE), 2011, vol., no., pp.40,46, 25-28 July 2011.

[9] Z.Z. Htike, "Can the future really be predicted?," Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2013 , pp.360-365, 26-28 Sept. 2013.

[10] Z.Z. Htike, "Multi-horizon ternary time series forecasting," Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), 2013, pp.337-342, 26-28 Sept. 2013.

[11] S.L. Win, et al., "Cancer Recurrence Prediction Using Machine Learning", International Journal of Computational Science and Information Technology (IJCSIT), 2014, Vol. 6, Issue 1.

[12] S.N.A. Hassan, et al., "Vision Based Entomology: A Survey", International Journal of Computer science and engineering Survey (IJCSES), 2014, Vol. 5, Issue 1, ISSN: 0976-2760.

[13] E.M. Azhari, et al., "Brain Tumor Detection and Localization in Magnetic Resonance Imaging", International Journal of Information Technology Convergence and services (IJITCS), 2014, Vol. 4, Issue 1, ISSN 2231-1939.

[14] N.A. Mohamad, et al., "Bacteria Identification from Microscopic Morphology: A Survey", International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), 2014, Vol. 3, Issue 1, ISSN: 2319-1015.

[15] S.L. Win, et al., "Cancer Classification from DNA Microarray Gene Expression Data Using Averaged One-Dependence Estimators", International Journal on Cybernetics & Informatics (IJCI), 2014, Vol. 3, Issue 1, ISSN: 2320-8430.

[16] E.M. Azhari, et al., "Tumor Detection in Medical Imaging: A Survey", International journal of Advanced Information technology (IJAIT), 2014, Vol. 4, Issue 1, ISSN 2231-1548.

[17] S.N.A. Hassan, et al., "Vision Based Entomology – How to Effectively Exploit Color and Shape Features", Computer Science & Engineering: An International Journal (CSEIJ), 2014, Vol. 4, Issue 1, ISSN: 2231-3583.

[18] S.L. Win, et al., "Gene Expression Mining for Predicting Survivability of Patients in Early Stages of Lung Cancer", International Journal on Bioinformatics & Biosciences, 2014, Vol. 4, Issue 1.