Predictive Analysis for the Arbovirus-Dengue using SVM Classification

A.Shameem Fathima[1] , D.Manimeglai[2]

[1]Research Scholar,Department of Computer Science and Engineering, Manonmanium Sundaranar University, Tirunelveli,India
[2]Department of Information Technology, National Engineering College, Tamilnadu, India

ABSTRACT

Data mining in biology and medicine is a core component of biomedical informatics, and one of the first intensive applications of computer science to this field. Today's biomedical data mining appears more multifaceted with advances in knowledge discovery in databases as well as machine learning approaches. This paper explores the application of machine learning technique- SVM for the identification of one of the Arboviral disease – Dengue. This paper reports novel biological discovery through nontrivial data mining process by using existing computational techniques. The goal of the system is to support the collection, and retrieval of public health documents, data, learning objects, and tools. We have deployed this generic infrastructure to facilitate data integration and knowledge sharing in the domain of dengue, which is one of the most prevalent viral diseases. This paper proposed an effort to apply the svm classification with the Radial basis function to classify the viral data and the model exhibits highly precise prediction rate.

Keywords: *Dengue, Machine learning, Support vector machine, Classification*

I. INTRODUCTION

Data mining (DM) is the intelligent computational analysis of large sets of data by using a combination of machine learning, statistical analysis and database technology, with the objective to discover patterns and rules useful for guiding decisions about future activities [1]. Current data mining tools are characterized by a surplus of algorithms but a lack of guidelines to select the precise method according to the nature of the problem under analysis. Learning at the base-level is focused on accumulating experience on a specific learning task e.g., medical diagnosis [2]. In the biological field, classification tools are used robustly in taxonomical classification to emulate the ambiguities in assigning the appropriate rank [1,2]. Vector-borne diseases often tend to be complicated and the enormous data generating from the experimental studies warrant a need for efficient tools, for analysis.

Although computer technology is undergoing a revolution its application in classification is still in its infancy, its potential, especially, has not been tapped in the arena of vector-borne diseases. Our goal is to conduct research in the identification of one of the Arbovirus - Dengue viruses. We have been interested in learning whether or not specific data for the dengue viruses are fittingly classifiable by using the theory of the machine learning namely using the SVM technique. The diagnosis and treatment of dengue is guided by the symptom and findings that the patient presents, and cannot depend on laboratory confirmation, since routine tests cannot confirm dengue with the speed required for patients in critical condition. The goal of data mining is predicting and generalizing a pattern to other data. It uses a variety of data analysis tools to discover patterns and relationships in data that can be used to make reasonably accurate predictions. It is a process, not a particular technique or algorithm. The rest of this paper is organized as follows: Section 1 provides an introduction to the topic of the research describing the problem that is discussed. Section 2 constitutes an overview of the literature survey done in various aspects of medical data mining. Section 3 describes the source of the viral data. Section 4 explains the methodologies and section 5 our proposed technique along with the limitations and constraints which the implementation imposes.

In Section 6 the experimental setup is presented and section 7 the results are discussed. Finally, in Section 8 overall conclusions have been conveyed along with perspectives of future work.

2. LITERATURE SURVEY

2.1. Problems in Medical Data Mining

Human medical data are at once the most rewarding and difficult of all biological data to extract and analyze. Extracting useful knowledge and providing scientific decision -making for the diagnosis and treatment of disease from the database increasingly becomes necessary. Data mining in medicine can deal with this problem [3].

2.1.1 Data Related Problems

The interest in systems for autonomous decision making in medical and engineering applications is growing, as data is becoming more easily available. Though the two areas medicine and engineering appear to be remote in terms of

the underlying processes, both face many common challenges. One of the problems of interest to both areas is autonomous prediction.

Because the medical information [4] is characteristic of redundancy, multi-attribution, incompletion and closely related with time, medical data mining differs from other one. The major areas of heterogeneity of medical data are:

1. Volume and complexity of medical data

2. Physician's interpretation

3. Sensitivity and specificity analysis

4. Poor mathematical characterization

5. Canonical form

Most pertinent works tend to focus on medical data heterogeneity. Simultaneous consideration of multiple data sources for the purposes of reasoning requires the development of highly complex data translation and unification methods – hence the attempts at developing and promoting standardization in medical data storage and processing [4]. The heterogeneity of medical data is, however, also expressed in the selection of semantics for individual data sets. Gauging and capitalizing on the importance of selected data pieces, such as X-ray and EMR images, ECG readouts and disease histories for the purposes of augmenting the diagnostic process proves a difficult task. All these problems call for new methods of analyzing multi-dimensional data. Yet, techniques such as association rule mining, clustering, outlier detection, supervised classification, social network analysis, time series analysis, text mining, etc., have been used with some degree of success. Data mining could indeed be applied on this data heterogeneous and very large to help on three different perspectives:(1) Patient record management to assist hospitals; (2) Decision support systems to aid medical practitioners and (3) Medical research to facilitate discoveries of outbreaks, new diseases, causes or remedies.

2.1.2 Tool Inadequacy

The analysis of medical data is typically aimed at constructing classification mechanisms, often expressed as decision trees or rule sets, to support qualified personnel in the decision making process. Unfortunately, due to the inadequacy of current knowledge extraction tools for solving medical problems, their results may prove dangerous to patients It is therefore imperative to verify automated diagnoses as well as the chains of reasoning which ultimately lead to particular conclusions.

2.2 Techniques involved in Medical Data Mining

Classification analysis is one of the widely adopted data mining techniques for healthcare applications to support medical diagnosis [5], improving quality of patient care, etc. If a training dataset contains irrelevant features (i.e., attributes), classification analysis may produce less accurate results. Feature selection is a preprocessing

technique commonly used on high-dimensional data and its purposes include reducing dimensionality, removing irrelevant and redundant features [6], reducing the amount of data needed for learning and improving algorithms' predictive accuracy. Feature selection is another problem that engineers face while mining medical data. Irrelevant attributes may bring noise to the data and harm the prediction abilities of generated models. On the other hand, eradicating important variables also may decrease the accuracy of the models. This is particularly important in case of the breast cancer data where a lot of factors can influence occurrence of the disease. Two methods are compared for selecting the most relevant features (variables) for medical data mining. The first one, data-driven is a mechanical approach utilizing automatic mechanisms. The second one is knowledge driven and based on opinions of experts. The experiments showed that feature selection done by the experts improves sensitivity of a classifier, while automated approach improves predictive power on the majority class.

Much research work in data mining is going on in improving the predictive accuracy of the classifiers by applying the techniques of machine learning such as SVM and feature selection [7]. The importance of feature selection in medical data mining is appreciable as the diagnosis of the disease could be done in this patient-care activity with minimum number of features. Feature selection [8] may provide us with the means to reduce the number of clinical measures made while still maintaining or even enhancing accuracy and reducing false negative rates. In medical diagnosis, reduction in false negative rate can, literally, be the difference between life and death.

Future developments in integrated medical data repositories, standardized data representation, and guidelines for the appropriate research use in medical is a growing volume of biomedical databases and repositories, the need to develop a set of tools to address their analysis and support knowledge discovery [8] is becoming acute. It is proposed to develop a substantial set of techniques for computational treatment of these data. The approaches in review are diverse in data mining methods and user interfaces and also demonstrate that the field and its tools are ready to be fully exploited in biomedical research.

3. SOURCES OF THE DIAGNOSTIC DATA

3.1 Problem Identification

The key element of a data mining study is to know what the study is for. Conducting a survey is often a useful way of finding something out, especially when `human factors are under investigation. Although surveys often investigate subjective issues, a well-designed survey should produce quantitative, rather than qualitative, results [1]. That is, the results should be expressed numerically, and be capable of rigorous analysis. At the first place, information and resources related to the study are collected by using the right techniques. This information must be obtained from the authorized and verified resources in order to fulfill the

standard and particular criterion. Next stage is to identify problem to be solved, aims and objectives and also the scope of the research. All this information may give an overview of the study and provide the idea to solve the problem.

3.2 Collection of Data

Professional literature broadly treats mining of medical data and proves that these techniques can be beneficial from the perspective of future diagnosis and treatment. This research is focused on developing a novel technique for predictive analysis applied to a real-life medical data set concerning the Dengue virus. The Dengue Program brought a lot of raw data in a form of surveys taken from different hospitals and diagnosis laboratories in Chennai and Tirunelveli from India. Each of them was filled out by both patients and doctors participating in the screening. The data contained in these surveys contain valuable information, encapsulated in various patterns and regularities, which is used in diagnosing process in the future. The data sets used in our experiments consists of 5000 samples with 29 symptoms associated with the disease. Each sample consists of few measurements with label that denotes its symptoms .Of these, some were dengue positive and some were Dengue negative though the symptoms seemed to be like dengue positive. Using a classifier for analysis of all clinical, hematological and virological data we obtained classification. The techniques can be used differently in different disease prevalence to yield clinically useful positive and negative predictive values.

3.3 Handling the Data

The first stage of the data mining process is to select the related data from many available databases to correctly describe a given task. There are at least three issues to be considered in the data selection. The first issue is to set up a concise and clear description of the problem. The second issue would be to identify the relevant data for the problem description. The third issue is that selected variables for the relevant data should be independent of each other[9].

4. BASIC METHODOLOGY IN USE

In general terms, Data Mining comprises techniques and algorithms, for determining interesting patterns from large datasets [1,7] There are currently hundreds (or even more) algorithms that perform tasks such as frequent pattern mining, clustering, and classification, among others. Understanding how these algorithms work and how to use them effectively is a continuous challenge faced by data mining analysts, researchers, and practitioners in particular because the algorithm behavior and patterns it provides may change significantly as a function of its parameters. In practice, most of the data mining literature is too abstract regarding the actual use of the algorithms and parameter tuning is usually a frustrating task. On the other hand, there are a large number of implementations available, such as those in the R project. The quality of data is more

related to algorithms and the separation of algorithms suitable for the dataset. Data can contain missing values; it means that not all attribute values are known. There are algorithms that operate with such data, others don't. There are also algorithms that only operate on nominal data, others on weighted. So the second step is to collect algorithms according to data specifics. The easiest way to collect algorithms is to choose a data mining environment.

5. PROPOSED METHODOLOGY

Putting a new novel technique to work in the scientific analysis of clinical data is a demanding task. It has to overcome various practical problems, from getting access to the right data over executing the right pre- and post-processing steps to make the analysis meaningful up to problems of finding the right computational resources. Our proposed technique that is constructed for an analysis platform supports the development and practical deployment of algorithms by facilitating their implementation and integration in existing workflows.

A major objective of this paper is to evaluate machine learning algorithms in medical and healthcare applications to develop a model that can help make timely and accurate decisions for Arboviral diseases especially the Dengue virus. In this paper, we use support vector machine (SVM) classifier for analyzing the dengue data .As far as we know, no prior study was conducted between this pioneer classification methods on a real-world medical information problem. The kernel and penalty parameters of the support vector machine (SVM) are determined by cross-validation. It is shown that SVM classifiers could improve the classification accuracy significantly. The ultimate dream, of course is to have some intelligent means available that can pre-process our data, apply the appropriate mathematical, statistical and artificial intelligence techniques, and then provide a solution and an explanation.

5.1 Support Vector Machines

Classifying data is a common task in machine learning. Suppose in some given data points each belong to one of two classes and the goal is to decide which class a new data point will be in. In the case of support vector machines, a data point is viewed as a p-dimensional vector (a list of p numbers), and we want to know whether we can separate such points with a $(p-1)$-dimensional hyper plane. This is called a linear classifier [10]. There are many hyper planes that might classify the data. One reasonable choice as the best hyper plane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyper plane so that the distance from it to the nearest data point on each side is maximized. If such a hyper plane exists, it is known as the maximum-margin hyper plane and the linear classifier it defines is known as a maximum margin classifier, or - equivalently - the perceptron of optimal stability. Support vector machines (SVMs) [11] are a set of

related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes the input is a member of, which makes the SVM a non-probabilistic binary linear classifier. Since an SVM is a classifier, then given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. Intuitively, an SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

This training algorithm [12], one of the most well-known of a class of performing methods for bio-data analysis uses the idea of kernel substitution for classifying the data in a high dimensional feature space. SVMs are based on the structural risk minimization principle, closely related to regularization theory. The goal is to find the separating plane with the largest margin (i.e., find the support vectors). Training a SVM is equivalent to solving a quadratic programming problem with linear constraints (the number of variables is equal to the number of training data). Two key elements in the implementation of SVM are the techniques of mathematical programming and kernel functions. Kernels can also be constructed to incorporate domain knowledge. This so-called 'kernel trick' gives the SVM great flexibility. The kernel we use is radial basis function. In the leave-one-out cross-validation, the disease status [13] in the data set is predicted while the rest of the data is regarded as the training set. In the leave-many-out cross-validation, n individuals are uniformly at random picked up from the data set, marked and put back, where n is the size of the data set.

6. EXPERIMENTAL SETUP

In this section, we describe our experiment setup, which are collected Dengue fever data from different hospitals and medical diagnosis labs in India. Based on this data we created an information table, and from the table we can generate the important symptoms for the dengue diagnosis.

6.1 Description of the Data set

Dengue is re-emerging throughout the tropical world, causing frequent recurrent epidemics. The initial clinical manifestation of dengue often is confused with other febrile states confounding both clinical management and disease surveillance. Evidence-based triage strategies that identify individuals likely to be in the early stages of dengue illness can direct patient stratification for clinical investigations, management, and virological surveillance. Here we report the identification of algorithms that differentiate dengue from other febrile illnesses in the primary care setting and predict severe disease in adults.

Dengue data was collected from over 5000 records of arboviral information specifically for dengue. Some were from patients who had dengue positive and others were having the symptoms of Dengue but found to be negative on doing the IgM and IgG tests. The following 29 attributes are real-valued input features of the patient record describing the symptoms:Age, Sex, fever, chills, Coryza , systolic, diastolic, Shock, Myalgia, Malaise, Arthralgia , Hallucinations, Confusion, Altered consciousness, Convulsion, Neck rigidity, Motor Weakness, paralysis, Lymphadenopathy, skin Rash, Symptoms, Pleural Effusion, Hb, RBC*103 cells/cu.mm, WBC, PC1, PC2, PC3 and the response variable report. These 29 features were used to predict the disease dengue.

6.2 SVM Usage in R project

For the experimental results, we are interested in the percentage of correctly classified instances of the algorithms (accuracy percentage) and the number of rules or size of trees produced by the classifiers. For the experimental setup, all the original medical datasets are entered in to excel sheet and saved as csv file format. Next, all the identified algorithms are tested to the medical dataset with the option of using 10-fold cross-validation and 100-cross validation. We report sensitivity, specificity, and accuracy of the prediction methods. We also report the risk rate of the corresponding integrated risk factor associated with each prediction method. It is computed as the ratio of the probability of developing disease among those predicted susceptible to the probability of developing disease among those predicted non-susceptible.

The R environment is an open source, multiple platform (e.g. Windows, Linux, Mac OS) and high-level matrix programming language for statistical and data analysis. R presents the advantage of being more flexible and extensible by design, thus integration of statistics, programming and graphics is more natural. Also, due to its open source availability and users' activity, novel DM methods are in general more quickly encoded into R than into commercial tools [14]. The R community is very active and new packages are being continuously created, with more than 2321 packages available at http://www.r-project.org/[15]. Thus, R can be viewed as worldwide gateway for sharing computational algorithms.

It addresses two important and common goals [14]

1. Classification { labeling a data item into one of several predefined classes; and
2. Regression {estimate a real-value (the dependent variable) from several (independent) input attributes. [14]

The R interface to libsvm in package e1071, and svm() [14] was designed to be as intuitive as possible. The R implementation is based on the S3 class mechanisms. It basically provides a training function with standard and formula interfaces, and a 'predict ()' method. In addition, a

'plot ()' method for visualizing data, support vectors, and decision boundaries is provided. Models are fitted and new data are predicted as usual, and both the vector/matrix and the formula interface are implemented. As expected for R's statistical functions, the engine tries to be smart about the mode to be chosen, using the dependent variable's type (y): if y is a factor, the engine switches to classification mode, otherwise, it behaves as a regression machine; if y is omitted, the engine assumes a novelty detection task.

The R implementation is based on the S3 class mechanisms. It basically provides a training function with standard and formula interfaces, and a predict () method. In addition, a plot () method for visualizing data, support vectors, and decision boundaries is provided. Hyper parameter tuning is done using the tune() framework, which performs a grid search over specified parameter ranges. We use a viral dataset to dengue diagnostic and apply SVM in it. The SVM model will be able to discriminate whether the patient has dengue or not. The first function is svm () [13] which is used to train a support vector machine. Some import parameters include:

data: an optional data frame containing the variables in the model.

type: sets how svm () will work. The possible values for classification are: C, nu

kernel: defines the kernel used in training and prediction. The options are: linear, polynomial, radial basis and sigmoid.

degree: parameter needed if the kernel is polynomial (default:3);

gamma: parameter needed for all types of kernels except linear (default: 1/(data dimension));

coef0: parameter needed for polynomial and sigmoid kernels (default: 0);

The steps involved in the technique are: [14]

1. Read the dataset

2. Prepare the Dataset

3. Divide at random the dataset in two subsets, one with about 70% of the instances to training, and another with around the remaining 30% of instances to testing

4. Choose the parameters: The range to gamma parameter is between 0.000001 and 0.1. For cost parameter the range is from 0.1 until 10.

5. Train the model

6. Test the model

Both for the SVM and the partitioning tree (via rpart ()), we fit the model and try to predict the test set values. Predictions from the model, as well as decision values from the binary classifiers, are obtained using the predict () method.

7. COMPUTATIONAL RESULTS

In this section we use a dataset of dengue viral diagnostic and apply svm in it. The analysis performed within this research are based on the data from the King Institute of Preventive Medicine and surveys filled out by patients and cards filled out by doctors from different hospitals. Data is extracted by using a standardized data collection form and is analyzed using R project version 2.12.2. The choice of the R project [13] as the computational platform stems from its popularity and thus critical mass, ease of programming, good performance, and an increasing use in several fields, such as bioinformatics and finances, among others. Correlations between mortality and symptoms, physical examination findings, or laboratory findings at admission are examined using logistic regression, where appropriate. Correlations between the development of respiratory failure and the aforementioned patient characteristics are examined similarly. The svm model will be able to discriminate dengue positive and negative cases. In this dataset there are 5000 instances and 29 attributes for each instance.

In this experiment, we applied evaluation methods including basic performance measures. These evaluation methods are based on the confusion matrix. The confusion matrix is a visualization tool commonly used to present performances of classifiers in classification tasks [14]. The intensity of effectiveness of the classification model is calculated with the number of correct and incorrect classifications in each possible value of the variables being classified in the confusion matrix.

The confusion matrix is used to compute true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN), as represented in Fig. 1.

		Predicted Class	
		Dengue Positive	Dengue Negative
Outcome	Dengue Positive	TP	FN
	Dengue Negative	FP	TN

Figure 1: Confusion Matrix

The analysis is done with the svm –classification using the e1071 package [16]. The data set is read and divided at random into two subsets, one with about 70% of the instances to training, and another with around the remaining 30% of instances to testing [14]. We use the C-classification with the radial kernel with two parameters cost and gamma. The task is to predict the type of a virus that has affected the patient based on the clinical symptoms of the patient. We use the "tune()" function to do a grid search over the supplied parameter ranges ("C" - cost, "\gamma" - gamma), using the train set. The range to gamma parameter is between 0.000001 and 0.1. For cost parameter the range is from 0.1 to 10. The parameters are

chosen and the model is trained and then tested. In 10-fold and 100-fold cross-validation tests on training, the predicted and the actual disease statuses are compared and the standard confusion matrix is obtained as shown in Fig 1.

The performance of a classification algorithm in data mining is greatly affected by the quality of data source. Each algorithm has its own advantages and disadvantages. There are three commonly used performance measurements including accuracy, sensitivity and specificity [14]. The accuracy of classifiers is the percentage of correctness of outcome among the test sets exploited in this study as defined in (7). The sensitivity is referred as the true positive rate, and the specificity as the true negative rate. Both sensitivity and specificity used for measuring the factors that affect the performance are presented in (8) and (9), respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{7}$$

$$\text{Sensitivity} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Specificity} = \frac{TN}{TN + FN} \tag{9}$$

Table 1. Performance Results of SVM

Quality Measure	SVM
Sensitivity	0.4723
Specificity	0.9759
Accuracy	0.9042
Risk Rate	0.8343

We also report the risk rate of the corresponding integrated risk factor associated with the prediction method. It is computed as the ratio of the probability of developing disease among those predicted susceptible to the probability of developing disease among those predicted non-susceptible.

8. CONCLUSION

Table 1 shows the performance of the machine learning technique implemented on the datasets. The best accuracy is achieved by svm - 0.9042. The good performance of SVM indicates that they can possibly be adjusted to improve specific ad hoc methods for prediction of susceptibility to complex diseases such as Dengue. The total accuracy for 10-fold and 100 cross-fold validation on training data for svm was 90.3 in both cases.

As shown in the experimental results, support vector machine has the highest classification precision most of the time .However support vector machine is very time consuming because of more parameters, and demands more computation time. This interdisciplinary research requires new tools and methods for searching, retrieving, manipulating and integrating data from multiple sources to generate, validate and apply new public health models. The resulting challenges are magnified in that the data and processes that we study are dynamic. When the system is developed, the flexibility of the design will make archival remotely sensed data invaluable for retrospective analyses of public health problems that have not been amenable to previous analysis. Our future work is the focus on mining with different types of arboviral diseases employing different machine learning techniques.

ACKNOWLEDGEMENTS

REFERENCES

[1]   J. Han and M. Kamber. Data Mining: Concepts and Techniques. MorganKaufmann, 2000.

[2]   Berry, M.J.A. and Linoff, G. (1997), Data Mining Techniques: For Marketing, Sales, and Customer Support, John Wiley, New York.

[3]   M. S. Chen, J. Han, and P. S. Yu. Data mining: An overview from a database perspective. IEEE Trans. Know ledge and Data Engineering, 8:866-883, 1996

[4]   Kononenko, I. Bratko, I., and Kokar, M. (1998), Application of machine learning to medical diagnosis, in Mi- chalski, RS, Bratko, and Kubat M. (Eds),Machine Learning in Data Mining: Methods and Applications,Wiley, New York, pp. 389-428

[5]   Fernando Alonso, Juan P. Caraca – Valente and Cesar Montes,“ Combining Expert Knowledge and Data Mining in a Medical diagnosis domain”, Expert Systems with application, pp. 367-375, Vol. 23, 2002.

[6]   Shusaku Tsumoto , Problems with Mining Medical Data,IEEE Trans 2000

[7]   Ian H. W., Eibe F. Data Mining: Practical Machine Learning Tools and Techniques. 2nd Edition– San Francisco: MorganKaufmann, 2005

[8]   Fernando Alonso, Juan P. Caraca – Valente and Cesar Montes,“ Combining Expert Knowledge

and Data Mining in a Medical diagnosis domain", Expert Systems with application, pp. 367-375, Vol. 23, 2002

[9] Kononenko, I. Bratko, I., and Kokar, M. (1998),Application of machine learning to medical diagnosis, in Mi- chalski, RS, Bratko, I and Kubat M. (Eds), Machine Learning in Data Mining: Methods and Applications,Wiley, New York, pp. 389-428.

[10] Mertik M., Kokol P., Zalar B.Gaining Features in Medicine Using Various Data-Mining Techniques //Computational Cybernetics ICCC 2005, IEEE rdInternational Conference. – 2005. – P. 21–24.

[11] N.Cristianini and J. Shawe-Taylor. An Introduction to Support Vector Machines.Cambridge University Press,2000

[12] The R Journal Vol. 2/1, June 2010

[13] www.r-project.org

[14] cran.r-project.org/web/packages/e1071/e1071.pdf

AUTHORS PROFILE

A.Shameem Fathima is a research student of Manonmanium Sundaranar University, India. She is persuing her Ph.d with specialization in computerscience and engineering under the supervision of Dr.D.Manimegalai. She has published her research results in leading international conference proceedings and journals

Dr.D.Manimegalai is currently the Head of Department of Information Technology, National engineering College,kovilpatti,India. She had her BE & ME from Government College of Technology, Coimbatore and PhD from Manonmaniam Sundaranar University, Tirunelveli She is working in National Engineering College in various positions. She has modest number of research publications including journals such as AMSE and Pattern Recognition letter and in National and International Conferences. Her current area of research interests include Medical Image Processing and Data Mining and Image Retrieval.

527