Diagnosis for Dengue Fever Using Spatial Data Mining

N.Subitha (M.phil, Research Scholar) [#1], Dr.A.Padmapriya M.C.A., M.phil., Ph.D.[*2],

Department of Computer Science & Engineering,
Alagappa University – Karaikudi, India.

***Abstract*— The research of spatial data is in its infancy stage and there is a need for an accurate method for rule mining. Association rule mining searches for interesting relationships among items in a given data set. This enable to extract pattern from spatial database using k-means algorithm which refers to patterns not explicitly stored in spatial databases. Since spatial Association mining needs to evaluate multiple spatial relationships among a large number of spatial objects. An interesting mining optimization method called progressive refinement can be adopted in spatial association analysis. The method first mines large data sets roughly using a fast algorithm and then improves the quality of mining in a pruned data set. The k-means algorithm randomly selects k number of objects, each of which initially represents a cluster mean or center. For each of the remaining objects, an object is assigned to the cluster to which it is most similar, based on the distance between the object and the cluster mean. Then it computes new mean for each cluster. This process iterates until the criterion function converges. The above concept is applied in the area of image segmentation where to apply the microscopic blood image as input and signals are filtered with the help of neural network to predict the best result about dengue fever.**

***Keywords*— Data mining, Spatial data mining, Spatial database, K-mean, Spatial relationship, Dengue fever.**

1. INTRODUCTION

Data mining is the process of discovering interesting, knowledge such as patterns, associations, changes, anomalies and significant structures, from large amount of data stored in database, data warehouses or other information repositories[2]. Due to the wide availability of huge amounts of data in electronic forms, the imminent need for turning such data into useful information and knowledge for broad application including market analysis, business management, and decision support, data mining has attracted a great deal of attention in information industry in recent years[3].

In general, a knowledge discovery process consists of an iterative sequence of the following step

1. Data cleaning,
2. Data integration,
3. Data selection,
4. Data transformation,
5. Data mining,
6. Pattern evaluation.
7. Knowledge presentation,

With the widely available relational database system and data warehouses, the four processes: can be performed by constructing data warehouses and performing some OLAP operations on the constructed data warehouses.

1.1 Data Mining Applications

Data mining is a young discipline with wide and diverse applications, there is still a nontrivial gab between general principles of data mining tools for particular applications.

1. Biomedical and DNA Data Analysis.
2. Financial Data Analysis.
3. Retail Industry.
4. Telecommunication Industry.

1.2 Data Mining Techniques

Data Mining Techniques Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbour method etc., are used for knowledge discovery from databases[5].

- Classification
- Prediction
- Association rule
- Neural networks

1.3 Spatial Data Mining

Spatial data mining refers to the extraction of knowledge, spatial relationships, or other interesting patterns not explicitly stored in spatial databases. Such mining demands an integration of data mining with spatial database technologies[2,4].

II. REVIEW OF RELATED WORKS

2.1 Motivation of the research work

Koperski, adhikary, et al had described about Spatial data mining, i.e mining knowledge from large amounts of spatial data, is a highly demanding field because huge amount of spatial data have been collected in various applications, ranging from remote sensing to geographical information system, computer cartography, environmental assessment and planning, etc. This paper summarizes recent works on spatial data mining from spatial data generalization, to spatial data clustering, mining spatial association rules, etc.

D.Rajesh suggested the concept is applied in the area of agriculture where giving the temperature and the rainfall as the initial spatial data and then by analyzing the agricultural meteorology for the enhancement of crop yields and also reduce the crop losses based on the k-means algorithm [2].

Raghuvira Pratap et al reviewed that mining knowledge from large amounts of spatial data is known as spatial data mining. It becomes a highly demanding field because huge amounts of spatial data have been collected in various applications ranging from geo-spatial data to bio-medical knowledge. The database can be clustered in many ways depending on the clustering algorithm employed, parameter settings used, and other factors. In this paper, an efficient density based k-medoids clustering algorithm has been proposed to overcome the drawbacks of DBSCAN and kmedoids clustering algorithms. The result will be an improved version of kmedoids clustering algorithm. This algorithm will perform better than DBSCAN while handling clusters of circularly distributed data points and slightly overlapped cluster.

Martin Ester, Hans-Peter Kriegel, Jorg Sander introduced a database-oriented framework for spatial data mining which is based on the concepts of neighbourhood graphs and paths.. Furthermore, techniques to efficiently support the database primitives by a commercial DBMS were presented. In the following sections, we covered the main tasks of spatial data mining: spatial clustering, spatial characterization, spatial trend detection and spatial classification. For each of these tasks, we presented algorithms as well as prototypical applications in domains such as the earth sciences and geography. Thus, we demonstrated the practical impact of these algorithms of spatial data mining.

Clustering or data grouping is one of the procedure in mining processing. This paper deals with the application of standard and k-means clustering algorithms in the area of image segmentation. In order to assess and compare both versions of k-means algorithm and its variants, appropriate procedures and software have been designed and implemented

2.2 spatial data mining

Spatial data mining is the process [8] of discovering interesting and previously un-known, but potentially useful patterns from large spatial

datasets. Extracting interesting and useful patterns from spatial datasets is more difficult than extracting the corresponding patterns from traditional numeric and categorical data due to the complexity of spatial data types, spatial relationships, and spatial autocorrelation. Spatial data mining , i.e., mining knowledge from large amounts of spatial data, is a highly demanding field because huge amounts of spatial data have been collected in various applications, ranging from remote sensing, to geographical information system(GIS), computer cartography, environmental assessment and planning[4] etc. It shows that spatial data mining is a promising field, with fruitful research results and many challenging issues.

Spatial Trend Analysis

Spatial trend analysis deals with the issue of detection of changes and trends along a spatial dimension. Typically, trend analysis detects changes with time, such as the changes of temporal patterns in time-series data. Spatial trend analysis replaces time with space and studies the trend of non-spatial or spatial data changing with space.

For example

The trend of changes in economic situation when moving away from the center of a city.

Applications of Spatial Data Mining

Some of the applications of spatial data mining are listed below,

- Geographic information systems,
- Geo marketing
- remote sensing
- image database exploration
- medical imaging
- navigation
- traffic control
- environmental studies

III. PROPOSED RESEARCH WORK

Introduction

Dengue Fever auscultation is the first basic analysis tool used to evaluate the functional state of the muscles, joints and bones. Microscope image test is ordered if the Blood from the Microscope slide shows any viruses or antibodies. Blood auscultation has the advantage of being viruses and seen on the screen, which gives a higher level of assurance of the accuracy of the initial diagnosis. However, because of viruses and antibodies, diagnosis may not be as accurate as desired. So we go for Microscope images report for automatic analysis of Dengue fever which involves following techniques classification, segmentation and feature extraction.

Classification

The classification is the process of splitting out the unwanted or background data

Here first the edge is detected before edge detection the smoothing and sharpening of image after the edge detection.

The intent of the classification process is to categorize all pixels in a digital image into one of several land cover classes, or *"themes"*. Normally, multispectral data are used to perform the classification and, indeed, the spectral pattern present within the data for each pixel is used as the numerical basis for categorization[6].

1. Supervised Classification
2. Unsupervised Classification

Edge detection

Definition of edges

Edges are significant local changes of intensity in an image.

Edges typically occur on the boundary between two different regions in an image.

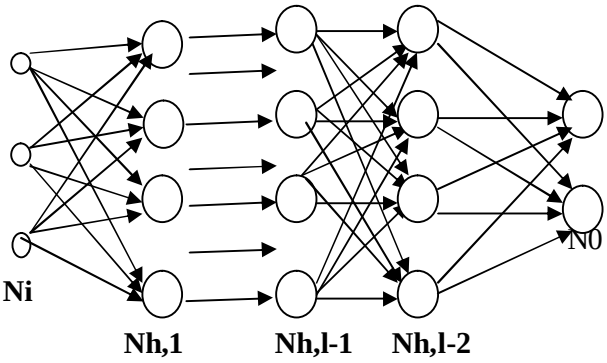
Segmentation

Image Segmentation is a subset of an expansive field of Computer Vision which deals with the analysis of the spatial content of an image[1]. In particular, it is used to separate regions from the rest of the image, in order to recognize them as objects. It is a method used in the vast field of Artificial Intelligence to identify the infected area.

Artificial Neural Networks

Artificial Neural Networks (ANNs) are a new approach that follows a different way from traditional computing methods to solve problems. Since conventional computers use algorithmic approach, if the specific steps that the computer needs to follow are not known, the computer cannot solve the problem. That means, traditional computing methods can only solve the problems that we have already understood and knew how to solve the critical problem during the short period.

Back Propagation ANNs contain one or more layers each of which are linked to the next layer. The first layer is called the "input layer" which meets the initial input and so does the last one "output layer" which usually holds the input's identifier. The layers between input and output layers are called "hidden layer(s)" which only propagate the previous layer's outputs to the next layer and [back] propagates the following layer's error to the previous layer



IV. EXPERIMENTAL METHODOLOGY

We use the following steps to implement the k-mean algorithm in the segmentation process.

Algorithm

K-Means Clustering

K-Means Algorithm: The algorithm for partitioning, where each cluster's center is represented by mean value of objects in the cluster.

Input:

k: the number of clusters,

D: a data set containing n objects.

Output: A set of k clusters.

Method:

1. arbitrarily choose k objects from D as the initial cluster centers.

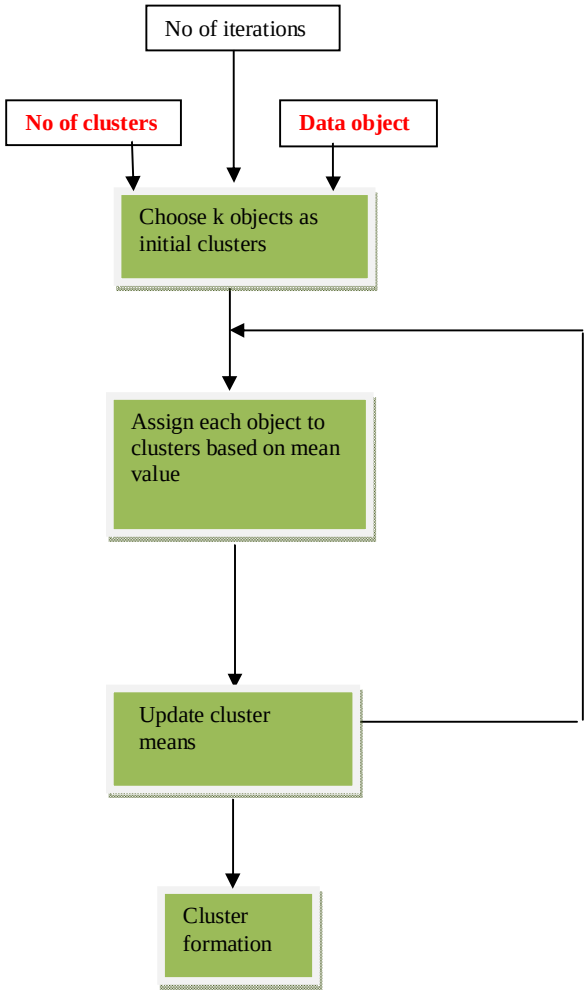
2. **repeat**

3. (re)assign each object to the cluster to which the object is most similar, based on the mean value of the objects in the cluster;.

4. update the cluster means;

45. **until** no alteration;

The initial assignment of points to clusters can be done randomly. In the course of the iterations, the algorithm tries to minimize the sum, over all groups, of the squared within group errors, which are the distances of the points to the respective group means. The groups obtained are such that they are geometrically as compact as possible around their respective means.



IV. IMPLEMENTATION RESULTS

We have to find dengue fever using the microscopic blood images are as follows

Figure 1. Layer management

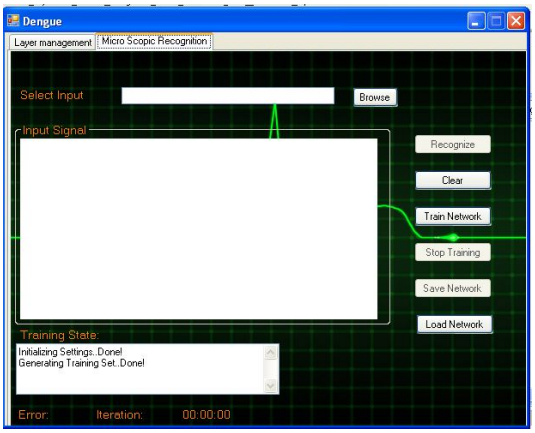


Figure 2. Process of train network

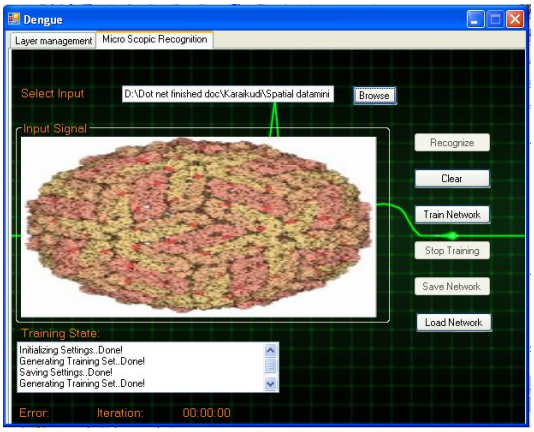


Figure 3. Virus recognition

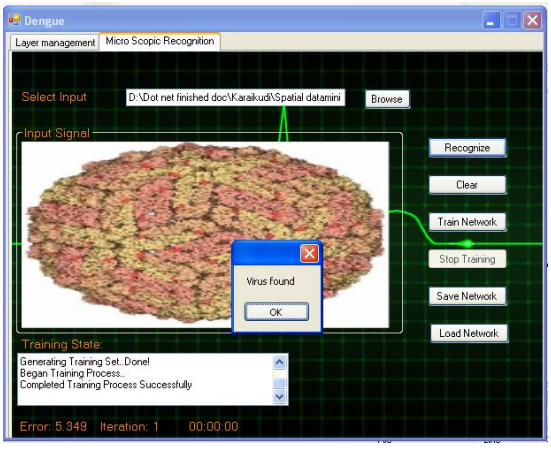


Figure 4. Virus detection

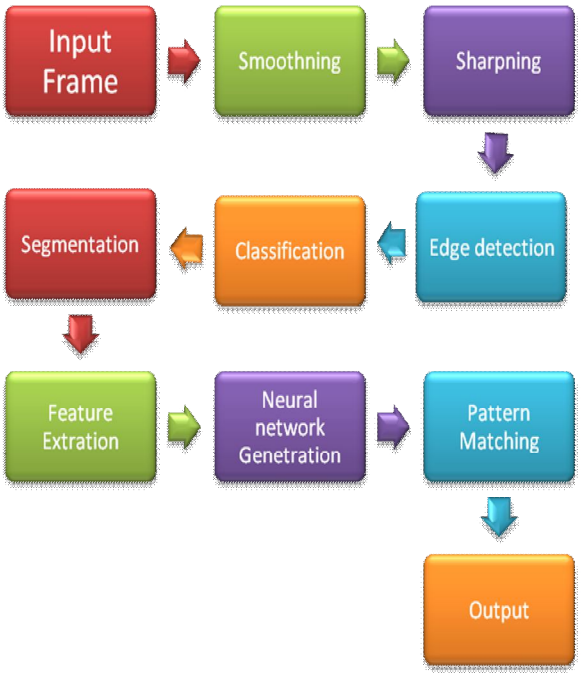


Figure 5. Execution flow of entire process

V. CONCLUSION

Dengue fever is transmitted by the bite of an ades mosquito infected with a dengue virus. In the past there are some automatic (i.e Dengue Test Kit) system to measure the Dengue fever if a patient want to know about the virus condition but it will take half an hour to predict the result sometimes prediction may be wrong.

In this paper we have implemented the automated system to reorganize the Dengue fever using the Microscope blood image report. Here we get Microscope image report as input and signals are filtered and the feature characteristics are extracted, the features are fed to the neural networks. Classification is carried using and the Back Propagation Network (BPN).Which gives the 98% accurate result with a short period.

REFERENCES

[1] Dariusz Małyszkoa, Sławomir T. Wierzchońb "Standard and Genetic *k*-means Clustering Techniques in Image Segmentation" Technical University of Białystok, Wiejska 45A, 15-351 Bialystok

[2] Ester M., Kriegel H.-P., and Sander J. 1997 "Spatial Data Mining: A Database Approach", Proc. 5th Int. Symp. on Large Spatial Databases, Berlin, Germany, pp. 47-66.

[3] U. M. Fayyades, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (Eds). Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press, 1996.

[4] W. Lu, J. Han, and B. C. Obi. Discovery of General Knowledge in Large Spatial Databases. In Proc. Far East Workshop on Geographic Information Systems pp. 275-289, Singapore, June 1993.

[5] 1. Jiawei Han and Micheline Kamber (2006), Data Mining Concepts and Techniques, published by Morgan Kauffman,

[6] Lillesand and Kiefer, 1994; Eastman, 1995"Image Classification

[7] S. B .Patil and Y . S. Kumaraswamy. " Intelligent and Effective dengue blood samples Prediction System Using Data Mining and Artificial Neural Network", European Journal of Scientific Research, 2009, ISSN 1450-216X Vol.31 No.4. pp.642-656 .

][4] Ramaswamy Palaniappan," Biological Signal Analysis", Ramaswamy Palaniappan and Ventus Publishing Aps, U.K, 2010

[8] Shekhar, S., and Chawla, S. 2003. Spatial Databases A Tour. Prentic e Hall (ISBN 0-7484-0064-6)

[9] P. Venkatesan* and S. Anitha. Application of a radial basis function neural network for diagnosis of diabetes mellitus. Current science, vol. 91, no. 9, 10 November, 2006, pp. 1195-1199.

10] F. Yaghouby, A. Ayatollahi and R. Soleimani, Classification ofCardiac Abnormalities Using Reduced Features of Heart Rate Variability Signal". World Applied Sciences Journal 6 (11), 2009, pp. 1547-1554.