International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 1, Issue 2, October 2014

A Food Prescription and Recognition for Patients using Decision Tree Algorithm

Kodeeswari.K[1], Sivananaitha Perumal.S[2]

PG Scholar, Department of IT, Dr.Sivanthi Aditanar College of Engineering, Tiruchendur, India[1]
Professor, Department of IT, Dr.Sivanthi Aditanar College of Engineering, Tiruchendur, India[2]

Abstract: The tendency for Data mining application in Healthcare Organizations is great because Healthcare sector is rich with information and Data mining is becoming a necessity. Healthcare organizations produce and collect large volumes of information on daily basis. Raw data in healthcare organizations are voluminous and heterogeneous. Computer vision-based food prescription with image could be used to estimate a meals and their calories content for diabetic patients. This study proposes a methodology for automatic food prescription, a visual dataset 50 food images was created and organized. The food and their calories are prescribed for the patients using decision tree algorithm. The collected data items are organized in Decision tree structure which can be used to predict the food for the particular disease. The food and calorie content will be prescribed based on the parameters such as age, gender and their severity. The system achieved classification accuracy of the order of 98%, thus proving the feasibility of the proposed approach is very challenging.

Keywords*:* Disease, BMI, Conceptual Clustering, Intersect method, Prediction.

I. INTRODUCTION

The treatment of Type 1 diabetic (T1D) patients involves exogenous insulin administration on a daily basis. A pran-dial insulin dose is delivered in order to compensate for the effect of a meal [1]. The estimation of the prandial dose is a complex and time-consuming task, dependent on many factors, with carbohydrate (CHO) counting being a key element. Clinical studies have shown that, in children and adolescents on inten-sive insulin therapy, an inaccuracy of±10 g in CHO counting does not impair postprandial control, while a±20 g variation significantly impacts postprandial glycaemia. There is also evidence that even well-trained T1D patients find it difficult to estimate CHO precisely [4]–[6]. In [4], 184 adult patients on intensive insulin were surveyed with respect to the CHO con-tent of their meals. On average, respondents overestimated the CHO contained in their breakfast by 8.5% and underestimated CHO for lunch by 28%, for dinner by 23%, and for snacks by 5%. In [5], only 23% of adolescent T1D patients estimated daily CHO within 10 g of the true amount, despite the selection of common meals. For children with T1D and their caregivers, a recent study has shown that 27% of meal estimations are inac-curate in ranges greater than±10 g.

Data mining is the process of extracting information from a data set and transforming into an understandable structure for further use and it is also the automatic extraction of useful, often previously unknown information from large databases or data sets. It is an increasingly important branch of computer science that examines data in order to find and describe patterns. Because we live in a world where we can be overwhelmed with information, it is imperative that we find ways to classify this input, to find the information we need, to illuminate structures, and to be able to draw conclusions. Computer assisted information retrieval may help support quality decision making and to avoid human error. Although human decision-making is often optimal, it is poor when there are huge amounts of data to be classified. Also efficiency and accuracy of decisions will decrease when humans are put into stress and immense work. Imagine a doctor who has to examine 5 patient records; he or she will go through them with ease. But if the number of records increases from 5 to 50 with a time constraint, it is almost certain that the accuracy with which the doctor delivers the results will not be as high as the ones obtained when he had only five records to be analyzed. Prediction algorithms determine models or rules to predict continuous or discrete target values given input data.

For example, a prediction problem could involve attempting to predict the amount of an insurance claim or a death rate given a set of inputs (picks one and then lists the corresponding inputs). Classification algorithms determine models to predict discrete values given input data. The increased number of diabetic patients worldwide, to-gether with their proven inability to assess their diet accurately raised the need to develop systems that will support T1D

patients during CHO counting. So far, a broad spectrum of mobile phone applications have been proposed in the literature, ranging from interactive diaries [7] to dietary monitoring based on on-body sensors [8]. The increasing processing power of the mobile de-vices, as well as the recent advances made in computer vision, permitted the introduction of image/video analysis-based appli-cations for diet management [9]–[14]. In a typical scenario, the user acquires an image of the upcoming meal using the camera of his phone. The image is processed—either locally or on the server side—in order to extract a series of features describing its visual properties. The extracted features are fed to a classifier to recognize the various food types of the acquired image, which will then be used for the CHO estimation.

A classification problem might involve trying to determine whether a particular purchase represents anomalous behavior based on some indicators (e.g., where the purchase was made, the amount of the purchase, or the type of purchase). Exploration uncovers dimensionality in input data. Trying to identify groups of similar customers based on spending habits for a large, targeted mailing is an exploration problem. Affinity analysis determines which events are likely to occur in conjunction with one another. A potential medical example would be analysis of patients' signs and symptoms that occur together in a clinical trial. Decision tree algorithm and Conceptual clustering algorithm are used in this project. Decision trees are a simple, but powerful form of multiple variable analyses. They provide unique capabilities to supplement, complement, and substitute for traditional statistical forms of analysis (such as multiple linear regressions), a variety of data mining tools and techniques (such as neural networks) and recently developed multidimensional forms of reporting and analysis found in the Field of business intelligence. Conceptual clustering algorithm is used to Finds clusters that share some common property or represent a particular concept.

II. Related Work

Measuring Calorie and Nutrition from Food Image [15] as people across the globe are becoming more interested in watching their weight, eating healthier, and avoiding obesity, a system that can measure calories and nutrition in every day meals can be very useful. To propose a food calorie and nutrition measurement system that can help patients and dietitians to measure and manage daily food intake. Our system is built on food image processing and uses nutritional fact tables. Recently, there has been an increase in the usage of personal mobile technology such as smart phones or tablets, which users carry with them practically all the time. Via a special calibration technique, our system uses the built-in camera of such mobile devices and records a photo of the food before and after eating it to measure the consumption of calorie and nutrient components. Our results show that the accuracy of our system is acceptable and it will greatly improve and facilitate current manual calorie measurement techniques.

The Use of Mobile Devices in Aiding Dietary Assessment and Evaluation [11] there is a growing concern about chronic diseases and other health problems related to diet including obesity and cancer. The need to accurately measure diet (what foods a person consumes) becomes imperative. Dietary intake provides valuable insights for mounting intervention programs for prevention of chronic diseases. Measuring accurate dietary intake is considered to be an open research problem in the nutrition and health fields. A novel mobile telephone food record that will provide an accurate account of daily food and nutrient intake. Our approach includes the use of image analysis tools for identification and quantification of food that is consumed at a meal. Images obtained before and after foods are eaten are used to estimate the amount and type of food consumed. The mobile device provides a unique vehicle for collecting dietary information that reduces the burden on respondents that are obtained using more classical approaches for dietary assessment. We describe our approach to image analysis that includes the segmentation of food items, features used to identify foods, a method for automatic por-tion estimation, and our overall system architecture for collecting the food intake information.

A Food Recognition System for Diabetic Patients Based on an Optimized Bag-of-Features Model [4] Computer vision-based food recognition could be used to estimate a meal's carbohydrate content for diabetic pa-tients. This study proposes a methodology for automatic food recog-nition, based on the bag-of-features (BoF) model. An extensive technical investigation was conducted for the identification and optimization of the best performing components involved in the BoF architecture, as well as the estimation of the corresponding parameters. For the design and evaluation of the prototype sys-tem, a visual dataset with nearly 5000 food images was created and organized into 11 classes. The optimized system computes dense local features, using the scale-invariant feature transform on the HSV color space, builds a visual dictionary of 10000 visual words by using the hierarchicalk-means clustering and finally classifies the food images with a linear support vector machine classifier. The system achieved classification accuracy of the order of 78%, thus proving the feasibility of the proposed approach in a very challenging image dataset.

III. System Analysis

A. Existing System

Doctors orally provide food prescriptions to patients in Healthcare Organizations. So they could miss some important food items that should be taken and also the patients could not remember all the food items .In such cases Doctors cannot handle more patients because it takes too much time. So, Doctors cannot maintain their planned Schedules.

Drawbacks

1. In earlier days, the food for patients had been given by oral.

2. By this, a doctor cannot handle more number of patients at a time.

3. Obviously, manual work increases and also patients cannot remember their food items.

B. Proposed System

Proposed System automatically provides food prescriptions in Healthcare Organizations by considering the patient's health conditions. The Dietician has to enter the characteristics such as Disease, Age, Gender, Height and weight. Based on the height and weight the BMI of the patients is calculated. Considering the calculated BMI the food is prescribed with appropriate Calories. It reduces the workload of Doctors.

Advantage

1. In this proposed system the food items will be provided to the patients in an automated system.
2. The automated system will reduce the doctors overhead.
3. More over the patients can easily follow \the food items prescribed by doctors by seeing the printed form.

C. Algorithm

a. Conceptual Clustering Algorithm

Data clustering is a method in which we make cluster of objects that are somehow similar in characteristics. The criterion for checking the similarity is implementation dependent. Clustering is often confused with classification, but there is some difference between the two. In classification the objects are assigned to pre defined classes, whereas in clustering the classes are also to be defined. Precisely, Data Clustering is a technique in which, the information that is logically similar is physically stored together. In order to increase the efficiency in the database systems the number of disk accesses is to be minimized. In clustering the objects of similar properties are placed in one class of objects.

Fig.1 Conceptual Cluster

Clustering is the process of partitioning data into a set of meaningful subclasses called clusters. Cluster is a collection of data items that are similar to one another and thus can be treated as a single group. Clustering is a unsupervised classification with no predefined classes. Conceptual clustering is one technique that forms concepts out of data incrementally by subdividing groups into subclasses iteratively; thus building a hierarchy of concepts. Conceptual Clustering is used to find clusters that share some common properties.

b. Decision Tree Algorithm

Decision trees are produced by algorithms that identify various ways of splitting a data set into branch-like segments. These segments form an inverted decision tree that originates with a root node at the top of the tree. The object of analysis is reflected in this root node as a simple, one-dimensional display in the decision tree interface. The display of this node reflects all the data set records, fields, and field values that are found in the object of analysis. The discovery of the decision rule to form the branches or segments underneath the root node is based on a method that extracts the relationship between the object of analysis (that serves as the target field in the data) and one or more fields that serve as input fields to create the branches or segments. The values in the input field are used to estimate the likely value in the target field. The target field is also called an outcome, response, or dependent field or variable. Once the relationship is extracted, then one or more decision rules can be derived that describe the relationships between inputs and targets. Rules can be selected and used to display the decision tree, which provides a means to visually examine and describe the tree-like network of relationships that characterize the input and target values.

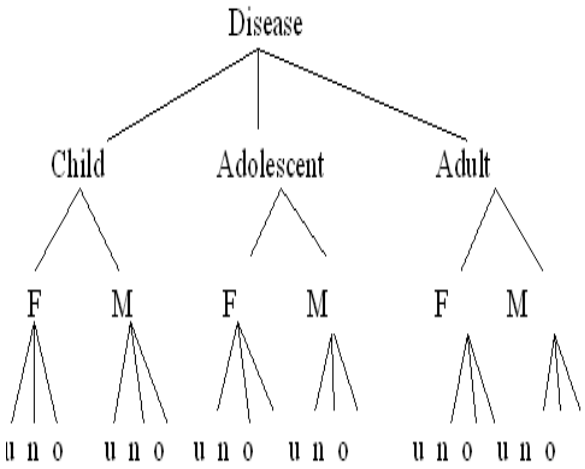
IV. Method Description

a. Decision tree organization of data items

Decision trees are a simple, but powerful form of multiple variable analyses. They provide unique capabilities to supplement, complement, and substitute for

a) a traditional statistical forms of analysis (such as multiple linear regression)
b) a variety of data mining tools and techniques (such as neural networks)
c) Recently developed multidimensional forms of reporting and analysis found in the field of business intelligence.

The collected data items from Healthcare Organization are organized in Decision tree structure. The characteristics of the patient are checked one by one in a hierarchical manner. Initially, the disease of the patient will be checked. Then the age of the patient will be checked and classified into a particular age group. Then gender will be checked and based upon the height and weight, BMI will be calculated. Finally, the leaf of the tree will be the table containing the appropriate food items.



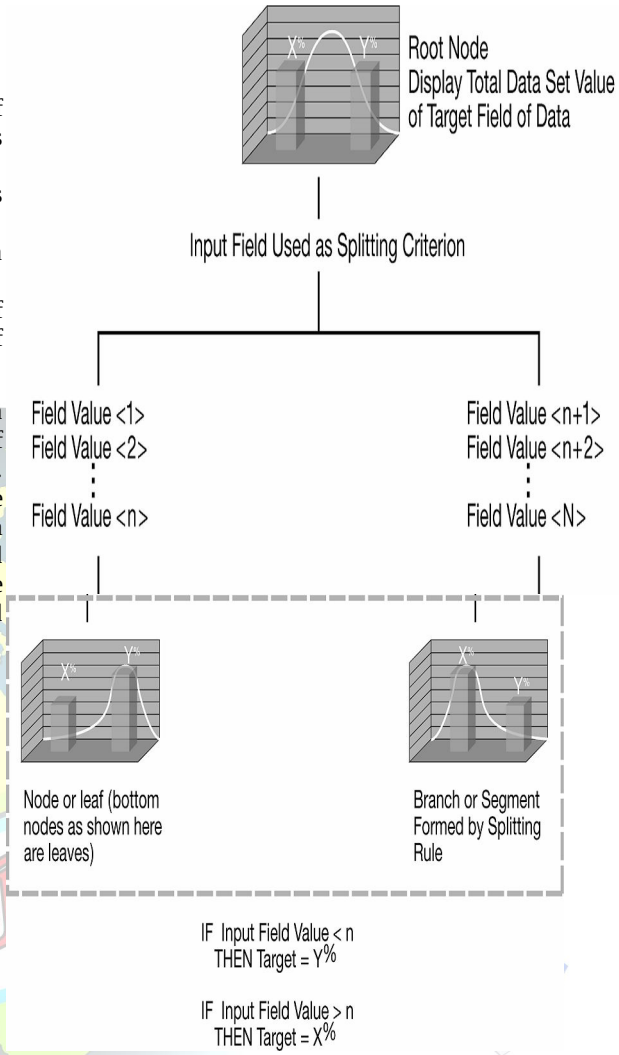


Fig.2 Decision Tree Algorithm

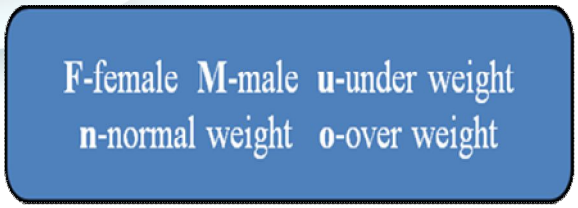


Fig.3 Decision Tree Structuring

b. Identification of major factors influencing health effects

Following are some of the common factors that influence health effects:

a) Disease
b) Age

c) Gender
d) BMI (Body Mass Index)

c. INTERSECT method of Conceptual Clustering

Conceptual clustering is one technique that forms concepts out of data incrementally by subdividing groups into subclasses iteratively; thus building a hierarchy of concepts. Conceptual Clustering is used to find clusters that share some common properties.

A new function, INTERSECT, is used to compare the similarity of two instances. Common attributes in two tables are selected based on INTERSECT method of Conceptual Clustering. INTERSECT method is used to find out the common food items. Food Prescription is provided based on the results obtained by Conceptual Clustering.

V. System Testing

Software testing is an investigation conducted to provide stakeholders with information about the quality of the product or service under test. Software testing can also provide an objective, independent view of the software to allow the business to appreciate and understand the risks of software implementation. Test techniques include, but are not limited to, the process of executing a program or application with the intent of finding software bugs (errors or other defects).Software testing can be stated as the process of validating and verifying that a computer program/application/product:

- meets the requirements that guided its design and development,
- works as expected,
- can be implemented with the same characteristics,
- satisfies the needs of stakeholders.

Software testing, depending on the testing method employed, can be implemented at any time in the development process. Traditionally most of the test effort occurs after the requirements have been defined and the coding process has been completed, but in the agile approaches most of the test effort is on-going. As such, the methodology of the test is governed by the chosen software development methodology.

Different software development models will focus the test effort at different points in the development process. Newer development models, such as Agile, often employ test-driven development and place an increased portion of the testing in the hands of the developer, before it reaches a formal team of testers. In a more traditional model, most of the test execution occurs after the requirements have been defined and the coding process has been completed.

VI. Results And Discussions

Figure 4 shows the Patient's details get from incoming diabetic patients such as patient's Name, Age, Gender, Height and Weight. Which are used to calculate the Body Mass Index of the Patient. Based on these criteria the diet plan is provided for the diabetic patient.

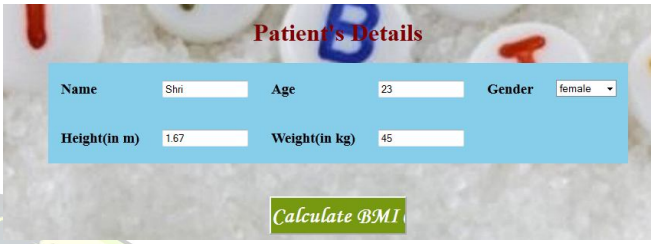


Fig.4 Patient's Details

Figure 5 shows the Body Mass Index of the patients.The BMI is below 18.5, they are under weight patients. The BMI is between the 18.5 to 24.9,they are the Normal weight patients. The BMI is Greater than 24.9, they are the Fat patients. Based on these the prescription is provided by the doctors

Fig.5 BMI Calculation

Figure 6 shows the Diet Plan for the diabetic patients such as Early Morning, Lunch and the Dinner is provided for the patients, also prescription and Note is provided.

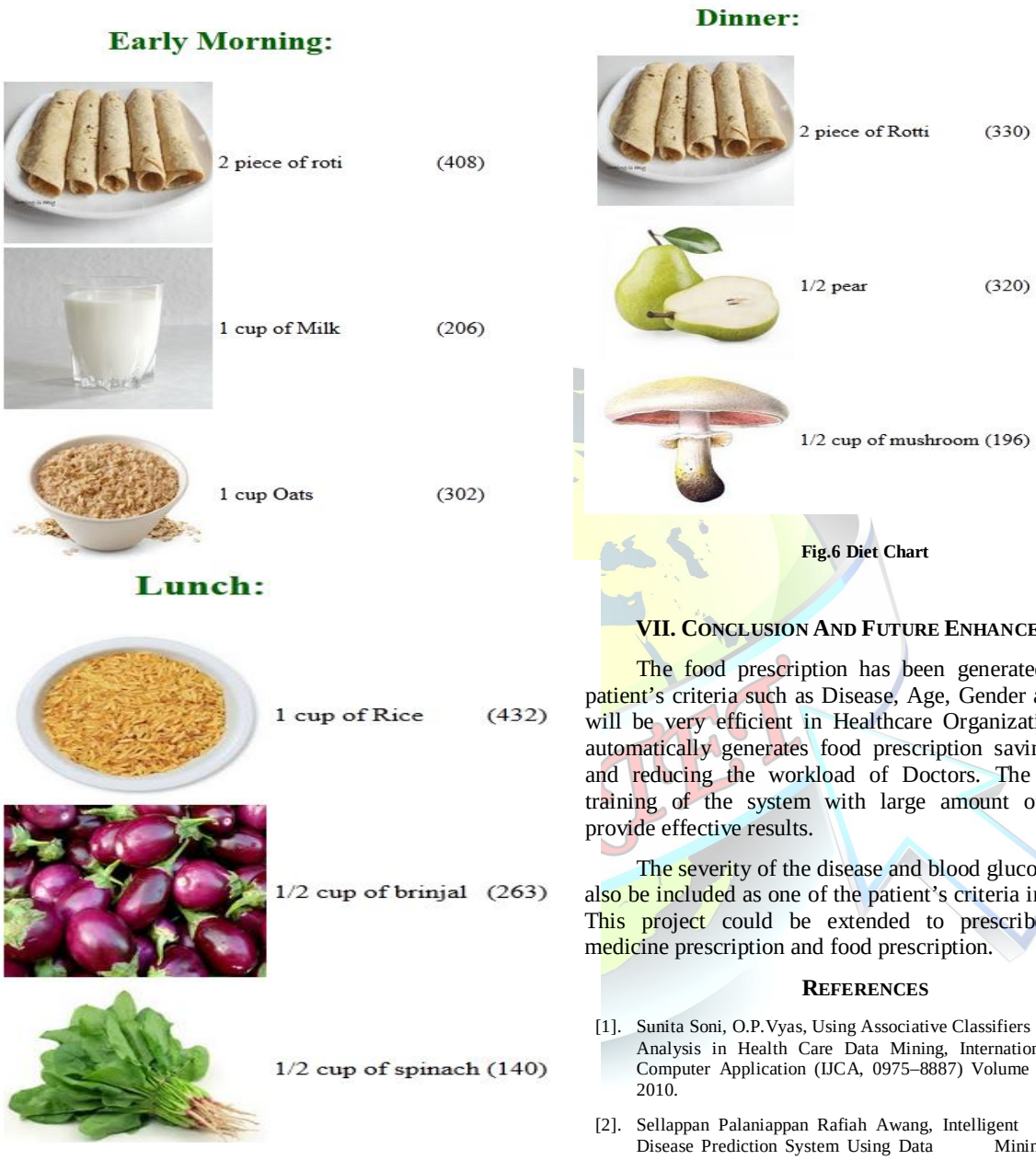


Fig.6 Diet Chart

VII. Conclusion And Future Enhancement

The food prescription has been generated based on patient's criteria such as Disease, Age, Gender and BMI .It will be very efficient in Healthcare Organization since it automatically generates food prescription saving the time and reducing the workload of Doctors. The continuous training of the system with large amount of data will provide effective results.

The severity of the disease and blood glucose level can also be included as one of the patient's criteria in the future. This project could be extended to prescribe for both medicine prescription and food prescription.

References

[1]. Sunita Soni, O.P.Vyas, Using Associative Classifiers for Predictive Analysis in Health Care Data Mining, International Journal of Computer Application (IJCA, 0975–8887) Volume 4– No.5, July 2010.

[2]. Sellappan Palaniappan Rafiah Awang, Intelligent IJCSNS Heart Disease Prediction System Using Data Mining Techniques, International Journal of Computer Science and Network Security, VOL.8 No.8, August 2008.

[3]. K.Park, Park's Textbook of Preventive and Social Medicine,Seventeenth Edition, Banarsidas Publications.

[4]. David G. Savage (2011-06-24). "Pharmaceutical Industr Supreme Court sides with pharmaceutical industry in two decisions". Los Angeles Times. Retrieved 2012-11-07.

[5]. Günnemann, Stephan; Kremer, Hardy; Seidl, Thomas (2011). "An extension of the PMML standard to subspace clustering models".

Proceedings of the 2011 workshop on Predictive markup language modeling – PMML '11.

[6]. Goebel, Michael; Gruenwald, Le (1999); A Survey of Data Mining and Knowledge Discovery Software Tools SIGKDD Explorations, Vol. 1, Issue 1.

[7]. Kobielus, James; the Forrester Wave: Predictive Analytics and Data Mining Solutions, Q1 2010, Forrester Research, 1 July 2008.

[8]. Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome (2009). "The Elements of Statistical Learning: Data Mining, Inference, and Prediction". Retrieved 2012.

[9]. Braha, Dan and Shmilovici, Armin (2003) On the Use of Decision Tree Induction for Discovery of Interactions in a Photolithographic Process, IEEE Transactions on Semiconductor Manufacturing.

[10]. Zhao, Kaidi; and Liu, Bing; Tirpark, Thomas M.;And Weimin, Xiao; A Visual Data Mining Framework for Convenient Identification of Useful Knowledge.

[11]. Mikut, Ralf; Reischl, Markus (September/October 2011)"Data Mining Tools". Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 1(5).

About Author

Kodeeswari.K received the B.Tech degree in Information Technology from Anna University, Chennai, in 2013. Currently, she is pursuing M.Tech degree in Information Technology from Anna University, Chennai. Her research areas of Interests include Data Mining, Operating system, and Networking.

Sivananaitha Perumal.S received the B.E degree in Electrical and Electronics Engineering from Anna University, Chennai, in 1996 and he received M.E in Power Systems from Anna University, Chennai, in Dec 1997. He pursued Doctrate in Electrical from Anna University, Chennai, in 2012. His research areas of Interests include Image Processing and Soft computing.