

A Review of Data Mining Classification Techniques Applied for Diagnosis and Prognosis of the Arbovirus-Dengue

A.Shameem Fathima¹, D.Manimegalai² and Nisar Hundewale³

¹ Department of Computerscience and Engineering , Manonmanium Sundaranar University ,
Tirunelveli, Tamilnadu, India

² Department of Information Technology, National Engineering College,
Kovilpatti, Tamilnadu, India

³ Department of Computerscience and Information Technology, Taif University, Saudi Arabia

Abstract

Chikungunya (CHIK) virus, similar to Dengue pose a serious threat in Tropics, because of the year-round presence of Aedes mosquito vectors. The use of machine learning techniques and data mining algorithms have taken a great role in the diagnosis and prognosis of many health diseases. But a very few work has been initialized in this arboviral medical informatics. Our focus is to observe clinical and physical diagnosis of chikungunya viral fever patients and its comparison with dengue viral fever. Our project aims to integrate different sources of information and to discover patterns of diagnosis, for predicting the viral infected patients and their results. The scope is mainly in the classification problem of these often confused arboviral infections. This study paper summarizes various review and technical articles on arboviral diagnosis and prognosis. In this paper we present an overview of the current research being carried out using the data mining techniques to enhance the arboviral disease diagnosis and prognosis. This paper is not intended to provide a comprehensive overview of medical data mining but rather describes some areas which seem to be important from our point of view for applying machine learning in medical diagnosis for our real viral dataset.

Keywords: Data Mining, Medical data, Machine learning algorithms, Diagnosis, Arbovirus.

1. Introduction

Presently, in most parts of the Tropics, epidemics are near peak transmission before they are recognized and confirmed as viral infection. By then it is too late to generally implement effective preventive measures that could have an effective impact on transmission and thus on the course of the epidemic. Therefore the surveillance for Dengue/ Chikungunya should be proactive. This proactive surveillance system will permit prediction of Dengue /

Chikungunya outbreak. The most important component of this system will also permit to differentiate whether the illness is Dengue or Chikungunya, as the initial symptoms are similar in both the disease. The objective of these predictions is to assign patients to either a “Dengue” group or a “Chikungunya” group or “any other infection” and to handle mystifying cases for the viral disease. Thus, arboviral diagnostic and prognostic problems are mainly in the scope of the widely discussed classification problems. These problems have paved a new face to many researchers in computational intelligence, data mining, and statistics fields.

Medical Informatics is generally clinical and/or biological in nature, and data driven statistical research has become a common complement. Predicting the outcome of a disease is one of the most interesting and challenging tasks where to develop data mining applications. As the use of computers powered with automated tools, large volumes of medical data are being collected and made available to the medical research group. As a result, Knowledge Discovery in Databases (KDD), which includes data mining techniques, has become a popular research tool for medical researchers to identify and exploit patterns and relationships among large number of variables, and made them able to predict the outcome of a disease using the historical cases stored within datasets. The objective of this study is to summarize various review and technical articles on diagnosis and prognosis of arboviral diseases. It gives an overview of the current research being carried out on various viral datasets using the data mining techniques to enhance the arboviral diagnosis and prognosis. 2. Arboviral Infections –An Overview of Dengue and Chikungunya

2. Arboviral Infections-An Overview of Dengue and Chikungunya

Chikungunya is a disease caused by the arbovirus that shares the same vector with dengue virus. Thus, in dengue-endemic region, Chikungunya is also a significant cause of viral fever causing outbreaks associated with severe morbidity. The symptoms of CHIKV infection are quite similar to those caused by many other infectious agents in the endemic areas. One particular difficulty in identifying CHIKV infection is its overlapping distribution with dengue viruses. It has been postulated that many cases of dengue virus infection are misdiagnosed and that the incidence of CHIKV infection is much higher than reported. Comprehensive study has not been undertaken to determine the clear picture of CHIKV infection and its comparison with dengue. Therefore, the present study was undertaken to diagnose Chikungunya infection in clinically suspected dengue patients and the vice-versa presented to The King Institute of Preventive Medicine, Chennai and in hospitals for diagnosis.

3. Knowledge Discovery and Data Mining

This section provides a preface to knowledge discovery and data mining. We provide the various analysis tasks that can be goals of a discovery process and lists methods and research areas that are challenging in solving these analysis tasks.

3.1. The Knowledge Discovery Process (KDD)

KDD is the process of extracting high-level knowledge from low-level data. Therefore, KDD refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data in databases. While data mining and KDD are often treated as the same words but in real data mining, it is an important step in the KDD process. The KDD process is often viewed as a multidisciplinary activity that encompasses techniques such as machine learning. The KDD process is interactive and iterative, involving numerous steps [1] such as

- (1) Data cleaning: also called data cleansing, is a phase in which noise data and irrelevant data are removed from the collection.
- (2) Data integration: is a stage in which heterogeneous multiple data sources are combined to a common source.
- (3) Data selection: at this step, the data related to the analysis is decided on and retrieved from the data collection.
- (4) Data transformation: also called as data consolidation, is a phase in which the selected data is transformed into forms appropriate for the mining procedure.

(5) Data mining: it is the crucial step in which knowledgeable techniques are applied to extract potentially useful patterns.

(6) Pattern evaluation: this step, strictly interesting patterns representing knowledge are recognized based on the given measures.

(7) Knowledge representation: is the final phase in which the discovered knowledge is visually represented to the user.

3.2. Data Mining Process

Data mining is the process of selecting, exploring and modeling large amounts of data in order to discover unknown patterns or relationships which provide a clear and useful result to the data analyst [2]. There are two types of data mining tasks: descriptive data mining tasks that describe the general properties of the existing data, and predictive data mining tasks that attempt to do predictions based on available data. In the context of the data mining tasks, diagnosis and prognosis are to discover knowledge necessary to interpret the gathered information. In some cases this knowledge is expressed as probabilistic relationships between clinical features and the proposed diagnosis or prognosis. In other cases, the system is designed as a black-box decision maker that is totally unconcerned with the interpretation of its decisions. Finally, in yet other cases, a rule-based representation is chosen to provide the physician with an explanation of the decision. The latest is the most convenient way for physician to express their knowledge in medical diagnosis. Thus, the major challenge presented by medicine is to develop technology to provide trusted hypotheses based on measures which can be relied upon in medical research and clinical hypothesis formulation.

Data mining involves some of the following key steps [3] -

- (1) Problem definition: The first step is to identify goals.
- (2) Data exploration: All data needs to be consolidated so that it can be treated consistently.
- (3) Data preparation: The purpose of this step is to clean and transform the data for more robust analysis.
- (4) Modeling: Based on the data and the desired outcomes, a data mining algorithm or combination of algorithms is selected for analysis. The specific algorithm is selected based on the particular objective to be achieved and the quality of the data to be analyzed.
- (5) Evaluation and Deployment: Based on the results of the data mining algorithms, an analysis is conducted to determine key conclusions from the analysis and create a series of recommendations for consideration.

4. Data Mining Classification Methods

Classification is the most frequently used data mining task with a majority of the implementation of Bayesian classifiers, neural networks, and SVMs (Support Vector Machines). A myriad of quantitative performance measures were proposed with a predominance of accuracy, sensitivity, specificity, and ROC curves. The latter are usually associated with qualitative evaluation.

Classification maps the data in to predefined targets. It is a supervised learning as targets are predefined. The aim of the classification is to build a classifier based on some cases with some attributes to describe the objects or one attribute to describe the group of the objects. Then, the classifier is used to predict the group attributes of new cases from the domain based on the values of other attributes. The commonly used methods for data mining classification tasks can be classified into the following groups [4].

4.1. Decision Trees (DT's)

A decision tree is a tree where each non-terminal node represents a test or decision on the considered data item. Selection of a certain branch depends upon the outcome of the test. To classify a particular data item, we start at the root node and follow the assertions down until we reach a terminal node (or leaf). A decision is made when a terminal node is approached. Decision trees that use recursive data partitioning can also be interpreted as a special form of a rule set, characterized by their hierarchical organization of rules.

4.2. Support Vector Machine (SVM)

Support vector machines (SVM) are based on statistical learning theory and belong to the class of kernel based methods. SVM is an algorithm that attempts to find a linear separator (hyper-plane) between the data points of two classes in multidimensional space. Such a hyper plane is called the optimal hyper plane. A set of instances that is closest to the optimal hyper plane is called a support vector. Finding the optimal hyper plane provides a linear classifier. SVMs are well suited to dealing with interactions among features and redundant features.

4.3. Genetic Algorithms (GAs) / Evolutionary Programming (EP)

Genetic algorithms and evolutionary programming are algorithmic optimization strategies that are inspired by the principles observed in natural evolution. Genetic

algorithms and evolutionary programming are used in data mining to formulate hypotheses about dependencies between variables, in the form of association rules or some other internal formalism.

4.4. Fuzzy Sets

Fuzzy sets form a key methodology for representing and processing uncertainty. Fuzzy sets constitute a powerful approach to deal not only with incomplete, noisy or imprecise data, but may also be helpful in developing uncertain models of the data that provide smarter and smoother performance than traditional systems.

4.5. Neural Networks

Artificial neural networks were recently the most popular artificial intelligence-based data modeling algorithm used in clinical medicine. Neural networks (NN) are those systems modeled based on the working of human brain. As the human brain consists of millions of neurons that are interconnected by synapses, a neural network is a set of connected input/output units in which each connection has a weight associated with it. The network learns in the learning phase by adjusting the weights so as to be able to predict the correct class label of the input. Neural networks may be able to model complex non-linear relationships, comprising an advantage over simpler modeling methods like the Naïve Bayesian classifier or logistic regression.

4.6. Rough Sets

The fundamental concept behind Rough Set Theory is similar to the Fuzzy set theory. The Difference is that the uncertain and imprecision in this approach is expressed by a boundary region of a set. Every subset defined through upper and lower approximation is known as Rough Set. Rough set is defined by topological operations called approximations, thus this definition also requires advanced mathematical concepts. They are usually combined with other methods such as rule induction, classification, or clustering methods.

5. Data Mining Classification Methods in use for the Data Mining of Arbovirus-Dengue

Clinical diagnosis of Dengue/ Chikungunya infection helps in predicting the viral cases. Suspected dengue case is defined as an acute febrile illness characterized by frontal headache, retro-ocular pain, muscle and joint pain, and rash (WHO, 2006). Besides the description of clinical symptoms, there are also clinical laboratory tests that are useful in the diagnosis of dengue. These clinical tests include a complete blood cell count (CBC), especially the

white blood cell count (WBC), platelet count and haematocrit levels. The results obtained from these methods are used to recognize the patterns which are aiming to help the doctors for classifying the malignant and benign cases. There are various data mining techniques, statistical methods and machine learning algorithms that are applied for this purpose. This section consists of the review of various technical and review articles on data mining techniques applied in arboviral dengue diagnosis.

In [5] Hani M. Aburas, B. Gultekin and Murat Sari predicted the dengue confirmed cases by using Artificial Neural Networks (ANNs). The model created by the authors were from 14,209 dengue reported confirmed-cases. They have taken many physical parameters such as mean temperature, mean relative humidity and total rainfall. Their prediction model has shown to be very effective processing systems for modeling and simulation in the dengue confirmed-cases data assessments as they did not use time information in building the model.

In [6] Janaína Gomide et al proposed a dengue surveillance approach that is a weekly overview of what is happening in each city compared with the weeks before. They construct a highly correlated linear regression model based on four dimensions: volume, location, time and content. Specifically, they showed that Twitter can be used to predict, spatially and temporally, dengue epidemics by means of clustering.

In [7], Silvia Rissino and Germano Lambert-Torres have used a Rough Set approach for the elimination of redundant data and the development of a set of rules that it can aid the doctor in the elaboration of the dengue diagnosis. From the dataset they roughly used for analysis they observed that patients with characteristics of all same attributes cannot be classified neither with dengue nor without dengue, but with only the decision attribute (dengue) not being identical and generates an inconclusive diagnosis for dengue.

In [8], Benjamin M. Althouse, Yih Yng Ng and Derek A. T. Cummings provided a comparison by analyzing dengue data from Singapore and Bangkok. Among the three models to predict incidence, SVM models outperformed logistic regression in predicting periods of high incidence. They found that the AUC for the SVM models using the 75th percentile cutoff is 0.906 in Singapore and 0.960 in Bangkok.

In [9], Ana Lisa V. Gomes et al, presented and implemented the novel application of the support vector

implemented the novel application of the support vector machines (SVM) algorithm to analyze the expression pattern of 12 genes in peripheral blood mononuclear cells (PBMCs) of 28 dengue patients (13 DHF and 15 DF) during acute viral infection. They achieved the highest accuracy of ~85% with leave-one-out cross-validation. However, their approach had a drawback that experimental investigation was necessary to validate their specific roles in dengue disease.

In [10], Fatimah Ibrahim, Mohd Nasir Taib, Wan Abu Bakar Wan Abas, Chan Chong Guan and Saadiah Sulaiman developed a prediction system based prediction solely on the clinical symptoms and signs. Their system uses the multilayer feed-forward neural networks (MFNN) and is able to predict the day of defervescence in dengue patients with 90% prediction accuracy.

In [11], Madhu.G, G.Suresh Reddy and Dr.C.Kiranmai presented an intelligent approach to dengue data analysis with rough sets for the elimination of redundant data and development of set of rules that can help medical practitioners in patient's diagnosis. They processed the data based on the lower and upper approximations and theory was defined as a pair of the two crisp sets to the approximations.

In [12], Sree Hari Rao and Suryanarayana U Murthy developed a novel efficient classification algorithm designated as VB Classif 1.0 which is utilized to classify nearly 10,000 records with 94% accuracy. This tool has performed better than the well known K-Nearest neighborhood (KNN) algorithm with different sizes of train data.

In [13], Fatimah Ibrahim, M. I Mohamad, S. N. Makhtar and J. Ibrahim, developed a rule based expert system to classify three types of risk-higher, lower and no risk group among the dengue infections using bioelectrical impedance analysis (BIA). The classification process was done according to gender, reactance value of the BIA and 'day of fever' on a daily basis diagnosis. Their system successfully classified the risk in dengue patients noninvasively with total classification accuracy of 66.7%.

In [14], F. Ibrahim, T. Faisal, M. I. Mohamad Salim and M. N. Taib, used bioelectrical impedance analysis (BIA) and artificial neural network (ANN) to analyze the data of nearly 223 healthy subjects and 207 hospitalized dengue patients. Four parameters were used for training and testing the ANN which are day of fever, reactance, gender, and risk group's quantification. Their Best ANN architecture trained with the steepest descent back propagation with momentum algorithm obtained the prediction risk

classification accuracy of 95.88% for high risk and 96.83% for low risk groups.

In [15], the authors aim to create a clinical Data warehouse for quick retrieval of reliable information on the viral diseases and their preventive measures at times of need. They claim that their ongoing work will be a boon to the researchers, academicians, Doctors, Health workers and Govt. servants and all for handy planning.

In [16], the authors employed decision tree algorithm to classify dengue infection levels into 4 groups (DF, DHF I, DHF II, and DHF III) and achieved an average accuracy is 96.50 %. The authors have compared their performance in term of false negative values to WHO and some researchers and found that that their research outperforms those criteria.

6. Proposed Work

There are many strains of Arbovirus. Our Research focus is on a particular variety. The main goal of the research is to first and foremost analyze the data from the surveys and to judge whether it is suitable to be analyzed with the use of the data mining methods. The second step is to evaluate several data mining algorithms in terms of their applicability to this data. Finally, an attempt is to be made to deliver some tangible medical knowledge extracted by the methods.

The analyses 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 2.12.0.

In summarizing we can separate the following methodological steps [17]:

1. Collecting and getting acquainted with a number of classification algorithms (e.g. data mining environment).
2. Reviewing the data set (e.g. a part of a patient health records).
3. Separating appropriate algorithms suitable for the data set.
4. Testing the full data set on selected number of classification algorithms, containing their default parameter values.
5. Selecting the best algorithms to use for further experiments.

6. Training the selected algorithms on reduced data set, by removing the attributes that appeared to be uninformative in building and visualizing the decision trees.

7. Modifying algorithms' default parameter values. Using the optimal data set formed for each algorithm of the most useful data identified in step 6.

8. Evaluating the results.

9. Randomizing the data set.

10. Performing steps 6 and 7 on randomized data set.

11. Evaluating and comparing results as well as algorithms performance.

These are the steps we have planned to perform with our data mining environment and data sets as well.

The choice of the R project [18] 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.

Several data mining models are built using decision trees, clusters, neural networks, logistic regressions, association rules and Naive Bayes. However, prior to this, the dataset has been pre-analyzed. This was to see how the attributes are represented in terms of their values to determine the initial input set of attributes. This was followed by the analyses.

The experiment within this research is to be conducted according to a defined formula. For each of the chosen data mining algorithms a set of models is built. Each of them is generated for a different parameters setting. These parameters vary from one method to another thus each of the models is treated individually. This means that one algorithm can have more models built than others.

The dataset has been split into two subsets: training and testing. The training dataset contains both discrete and continuous attributes. Some of the algorithms in the R project require the input attributes to be discrete only, though. Thus the models have been divided into two groups with respect to this facet. The first group contains those that accept continuous attributes: decision trees,

clustering, neural networks and logistic regression. The other comprises algorithms which require the attributes to be discrete: association rules and Naive Bayes. After the models have been constructed, they undergo the evaluation step. Their performance is measured with the use of lift charts and classification matrices. Afterwards, from each of the methods the best model (the best parameters setting) is chosen for the ultimate comparison of the algorithms. Finally, the best model emerges. [19][20]

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 data invaluable [21] for retrospective analyses of public health problems that have not been amenable to previous analyses.

7. Conclusion

The presented discussion on knowledge extraction from medical databases is merely a short summary of the ongoing efforts in this area. It does, however, point to interesting directions of our research, where the aim is to apply hybrid classification schemes and create data mining tools well suited to the crucial demands of medical diagnostic systems. It is proposed to develop a substantial set of techniques for computational treatment of these data. The approaches in review [22] 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.

Acknowledgments

A heartfelt gratitude is expressed by the authors of this paper to the Department of Virology, King Institute of Preventive Medicine and Research, Chennai, India, for releasing the viral data for research and education.

References

[1] Fayyad, U. M.; Piatetsky-Shapiro, G.; Smyth, P.; and Uthurusamy, R. 1996. *Advances in Knowledge Discovery and Data Mining*. Menlo Park, Calif.: AAAI Press.
[2] P. Giudici, *Applied Data Mining Statistical Methods for Business and Industry*, Wiley & Sons, 2003.
[3] *The Data Mining Process*. [Online].
http://publib.boulder.ibm.com/infocenter/db2luw/v9/index.js?topic=/com.ibm.im.easy.doc/c_dm_process.html.

[4] Han J. and Kamber M., *Data Mining: Concepts and Techniques*, 2nd ed., San Francisco, Morgan Kaufmann Publishers, 2001.
[5] Hani M. Aburas, B. Gultekin Cetiner and Murat Sari, Dengue confirmed-cases prediction: A neural network model, *Expert Systems with Applications: An International Journal*, Volume 37 Issue 6, June, 2010
[6] Janaína Gomide, Adriano Veloso, Wagner Meira Jr., Virgílio Almeida, Fabrício Benevenuto, Fernanda Ferraz and Mauro Teixeira Dengue surveillance based on a computational model of spatio temporal locality of Twitter, *Journal Web science 2011 ACM*.
[7] Silvia Rissino and Germano Lambert-Torres, *Rough Set Theory – Fundamental Concepts, Principles, Data Extraction, and Applications*, *Data Mining and Knowledge Discovery in Real Life Applications*, pp. 438, February 2009
[8] Benjamin M. Althouse, Yih Yng Ng and Derek A. T. Cummings, Prediction of Dengue Incidence Using Search Query Surveillance, *PLoS Negl Trop Dis*. 2011 August; 5(8): e1258.
[9] Ana Lisa V. Gomes, Lawrence J. K. Wee, Asif M. Khan, Laura H. V. G. Gil, Ernesto T. A. Marques, Jr, Carlos E. Calzavara-Silva and Tin Wee Tan, Classification of Dengue Fever Patients based On Gene Expression Data Using Support Vector Machines, *PLoS One*. 2010; 5(6): e11267
[10] Fatimah Ibrahim, Mohd Nasir Taib, Wan Abu Bakar Wan Abas, Chan Chong Guan, Saadiah Sulaiman, A novel dengue fever (DF) and dengue haemorrhagic fever (DHF) analysis using artificial neural network (ANN), *Computer Methods and Programs in Biomedicine*, Volume 79 Issue 3, September, 2005, p273-281
[11] Madhu. G, G. Suresh Reddy and Dr. C. Kiranmai, Hypothetical Description for Intelligent Data Mining, *International Journal on Computer Science and Engineering*, Vol. 02, No. 07, 2010, 2349-2352
[12] Sree hari Rao Vadrevu, Suryanaryana U Murthy, A Novel Tool For Classification of Epidemiological Data of Vector Borne Diseases, *Journal Of Global Infectious Diseases*, Jan-Apr 2010.
[13] Fatimah Ibrahim, M. I. Mohamad, S. N. Makhtar and J. Ibrahim Classification of Risk in Dengue Fever and Dengue Haemorrhagic Fever using Rule Based Expert System, *IFMBE Proceedings*, 2007, Volume 15, Part 3, 50-53.
[14] F. Ibrahim, T. Faisal, M. I. Mohamad Salim and M. N. Taib, Non-invasive diagnosis of risk in dengue patients using bioelectrical impedance analysis and artificial neural network, *Medical and biological engineering and computing* Volume 48, Number 11, 1141-1148.
[15] Dr. M. Usha Rani, M. Kalpana Devi, D. M. Mamatha, R. Seshadri, Yaswanth Kumar. Avulapati, Clinical Data Warehouse on Insect Vector Diseases to Human of Andhra Pradesh, *International Journal of Computer Science and Information Security*, 2010, Vol 8, Issue 5, P 240-244.
[16] Daranee Thitiprayoonwongse, Prapat Suriyaphol, and Nuanwan Soonthornphisaj., Data Mining on Dengue Virus Disease, 13th International Conference on Enterprise Information Systems (ICEIS 2011), pp. 32-41, June 8-11, 2011

- [17] Mertik M., Kokol P., Zalar B. Gaining Features in Medicine Using Various Data Mining Techniques // Computational Cybernetics ICC C 2005, IEEErd International Conference. – 2005. P. 21–24
- [18] www.r-project.org.
- [19] Rodríguez, A., Carazo, J.M. and A., Trelles O.; “Mining Association Rules from Biological Databases”; Journal of the American Society for Information Science and Technology; 56(493-504); 2005
- [20] S. Sfakianakis, M. Blazantonakis, I. Dimou, M. Zervakis, M. Tsiknakis, G. Potamias, D. Kafetzopoulos, D. Lowe, Decision Support Based on Genomics: Integration of Data and Knowledge Driven Reasoning, International Journal of Biomedical Engineering and Technology, Special Issue on Decision Support in Medicine.
- [21] Quinlan, J. R. (1983). Learning efficient classification procedures. In Machine Learning: An Artificial Intelligence Approach, ed. Michalski, Carbonnel & Mitchell. Tioga Press.
- [22] Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann.

A. Shameem Fathima is currently a Ph.D student in the Department of Computer Science and Engineering at Manonmanium Sundaranar University, India. She obtained M.E in Computer Science and Engineering from Crescent Engineering College affiliated to Anna University in 2004. She has 5 years of teaching experience in different academic institutions in India and abroad. She has a proven career record and has published many papers in conferences. Her focus of research is Data Mining.

Dr. D. Manimegalai is Professor and Head of the Department of Information Technology in National Engineering College. She had her BE & ME from Government College of Technology, Coimbatore and PhD from Manonmaniam Sundaranar University, Tirunelveli. Her Current areas of research interests include Medical Image Processing and Data Mining and Image Retrieval. She is a life member of Computer Society of India, Institution of Engineers, System Society of India and Indian Society for Technical Education

Dr. Nisar Hundewale received his Ph.D. in Computer Science from Georgia State University, USA. He has worked at National Institutes of Health (NIH), USA, as post-doctoral Fellow. Currently, he is an Assistant Professor and Associate Dean for Research at Taif University. His research interests are Algorithms, Machine Learning, Bioinformatics, Distributed Computing, and Networking. He is a great inspiration and shore up to young researchers.