International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 12 | Dec - 2021 ISSN: 2582-3930

Heart Disease Identification using Hybrid Classification

Ashima (Research scholar) , Kiranpreet Kaur (Assistant Professor)

ABSTRACT

Data mining, a method utilizes for mining the valuable data from the rough data. The futuristic outcomes are forecasted using recent information in the prediction analysis. This research work deals with the prediction of the heart disease. There are several steps that are included in the heart disease prediction. The RF and DT based hybrid scheme is introduced. The features are abstracted using RF. The implementation of DT is carried out for classification. The performance analysis of the recommended model for acquiring accuracy, precision and recall is completed in this research. The accuracy has obtained in predicting the heart disease from this model is evaluated 94.44%.

KEYWORDS

Heart Disease, DT, RF, Machine learning

1. INTRODUCTION

Heart disease defines any illness related to heart. The problems using the blood vessels, cardiovascular system and the heart are defined as the cardiovascular disease [9]. On the other hand, the problems and deformities in the heart itself are known as heart disorder. A major cause of deaths in the United Kingdom, US and Australia is cardiac ailment as stated by the CDC. The heart disease may lead to every one out of deaths in Union States. The heart disease consists of numerous kinds of diseases due to which various parts of the organ are infected. The Congenital heart disease, Arrhythmia, CAD, Myocardial infarction, Heart failure, Hypertrophic cardiomyopathy, pulmonary stenosis etc. is the major types of heart diseases. The Arrhythmia is a kind of the heart rhythm abnormality. The plaque is grown in the arteries of heart due to CAD that is also called ischemic heart disease. The other name of Myocardial infarction is heart stroke.

The signs of this disorder are occurred according to the condition that affected the individual. But, chest discomfort, breathing difficulty etc are some main signs of this disorder. The common symptom that can be seen in various types of heart diseases namely angina, or angina pectoris is chest pain. These diseases happen due to lack of enough oxygen in heart. The stressful events or physical exertion may lead to Angina and it lasts under ten minutes. When harm occurs in a portion or whole organ, in the coronary arteries or when nutrients and oxygen are not supplied properly to the organ, the heart disease occurs. There

are some genetic heart diseases named hypertrophic cardiomyopathy. These diseases occurred in a person before the person is born besides inborn heart defects [10]. The severity level of this disorder can be maximized due to numerous choices of lifestyle. High BP and fat, smoking, bad food habits, oldness, sitting too long at work are some habits that can cause heart attack. The use of data mining technology for the healthcare sector is revolutionized the task of disease prediction. The role of this technology in heart disease prediction is quite significant. At present, various data mining methods are being carried out for detecting and extracting valuable knowledge from the medical data set using minimum user inputs and hard work. With the time, researchers have explored several methods for implementing data mining in medical domain so that diverse kinds of cardiovascular ailments can be forecasted with accuracy. The performance of data mining differs from technique to technique being adopted and chosen attributes. In general, the clinical databanks in the medical domain are useless and unpredictable.

The different steps shown in the above heart disease prediction model have been described below:

a. Data pre-processing: This is the initial step while forecasting of heart disorder. In this step, a number of pre-processing methods are applied on the input data for introducing completeness. This is done to perform meaningful analysis of the data and getting optimal outcomes. In data cleaning process, the detection and removal of repetitive archives, spelling mistakes and unlikely data is carried out. Similarly, missing data and outliers are addressed. Initially, the marking of the missing values in the data is carried out by means of a mathematical filtering approach. This filter does the cleaning of the numeric data, which may be too large or too small [11]. The filtering approach set this data to a defined default value. Once the missing values are marked and detected, the used fileting approach replaces these missing values using mean value of the allocated data. The non-redundant and denoise data is given as input to the feature selection process. In this step, redundant features are eradicated from the dataset. This may make the performance of training model better.

b. Feature selection: In heart disorder prediction, the step of feature selection focuses on the selection of subset of extremely unique features. In this step, those unique features are selected which relate to different existing classes. The dataset of heart disease may contain a number of features. However, merely a few of these features contribute significantly in the decision-making process to classify diseases. This is the reason that the feature selection is implemented on the dataset values so that it can reduce the size of feature vector to an expressive sample size. There are mainly two tasks included in the feature selection process. At first, a feature evaluator method is implemented for evaluating the dataset features in terms of the output class. In the second task, a search algorithm is applied. This algorithm uses different

mixtures of features for selecting most appropriate set to handle classification issue.

c. Prediction: In this step, the mapping of selected features is carried out onto the training model for classifying the given features. In order to generate predictions, a specialist cardiologist does the labelling of gathered heart disease dataset. The classification is developed as a multi-class issue and the classification of medical data is carried out into classes of 4 types. Hence, every class refers to a certain category of heart disease [12]. This task can find the probabilities of disorder in patients based on the chosen crucial attributes.

2. LITERATURE REVIEW

Mamatha Alex P and Shaicy P Shaji,et.al (2019)studied that many forms of heart disease can be detected or diagnosed with different medical tests by considering family medical history and other factors. But, the prediction of heart diseases without doing any medical tests is quite difficult. The aim of this project is to diagnose different heart diseases and to make all possible precautions to prevent at early stage itself with affordable rate. We follow 'Data mining' technique in which attributes are fed in to SVM, Random forest, KNN, and ANN classification Algorithms for the prediction of heart diseases. The preliminary readings and studies obtained from this technique is used to know the possibility of detecting heart diseases at early stage and can be completely cured by proper diagnosis. There was a list included thirteen features that had affected the probability of the person who was suffered from the heart disease had recognized in this paper.

Imran Mirza, et.al (2019) focused on the data science arena to perform the scrutiny of clinical sector in which the significant insight and predictions were offered that assisted the physicians globally [24]. For this purpose, the implementation of machine learning algorithms was recommended for performing data mining on clinical database. The patterns in the data were investigated in this way and predictions were generated accurately in.

Anjan Nikhil Repaka, et.al (2019) emphasized on the detection of heart disease in which earlier data and information had considered [16]. For this purpose, the Navies Bayesian was utilized to develop the SHDP that was capable for risk factors prediction related to heart disease. The significant growth in mobile healthcare had seen due to rapid advancement of technology. The required data was brought together in a standardized form. The extraction of some features had performed for predicting the possibilities of heart disorders in the person who suffered with this disease. Data mining algorithms were utilized for the presentation and elaboration of multiple knowledge abstraction schemes.

Cincy Raju, et.al (2018) examined that the most dangerous disease was heart disease which caused the death. This disease also caused a severe long-term disability [21]. This disease attacked on human being very quickly. Thus, the diagnosis of patient was required with accuracy in terms of time as it was very demanding function for medical support. The hospitals whose diagnosis was proved invalid lost their reputation. One of the leading biomedical problems was accurate analysis of heart disease. The data mining methods were carried out for building an efficient treatment in which curative situations were facilitated.

Aakash Chauhan, et.al (2018) analyzed a significant cause of short life was heart disorder in the contemporary society [22]. There were a lot of people who relied on the healthcare system for acquiring result with accuracy in less time. The healthcare organization had generated and gathered the massive data each day. The extraction of data was facilitated in the data innovation using atomization of processes so as the interesting knowledge was acquired.

Rashmi G Saboji, et.al (2017) recommended a scalable framework to predict health related disorders with the help of certain features [19]. The healthcare data was utilized to achieve this. The major purpose of this work was that the heart disease detection was predicted using few features. The random forest was implemented on Apache Spark in this prediction solution..

M. A. Jabbar, et.al (2016) evaluated that the coronary heart disorder was one among the main causes of fatalities globally [23]. Heart disorder discovery was very wearisome task. An intelligent decision support system had required for forecasting the disease. A person being disease or not was classified using Data mining methods. The HNB was a data mining model in which the conditional independence assumption of the conventional Naïve Bayes was settled down. This suggested model asserted that with the implementation of HNB, the heart disease was successfully predicted.

Ankita Dewan, et.al (2015) suggested acompetent genetic algorithm along with BPapproach to predict health related disorders [17]. The unknown knowledge of heart illness was determined and extracted from a historic database by prototype that was the fundamental goal of this paper. This technique had potential to deal with the complicated queries to predict health related disorders.

Monika Gandhi, et.al (2015) analyzed that plenty of information that had generated in the medical associations, still this information was not exploited in proper manner [18]. There was abundance of data in health care system but the knowledge was very poor. The successful analysis techniques were not present for determining associations and patterns in clinical data. The data mining techniques were able and proved as remedy in this situation. Thus, the deployment of data mining techniques had done for this. This paper proposed to provide details related to different approaches of info retrieval through data mining

techniques. These techniques were carried to predict health related disorders in research.

T. John Peter, et.al (2012) suggested pattern recognition and data mining algorithms. These algorithms were carried out in medical risk forecasting models [20]. The classification data mining method was utilized to model and categorize the data. The basic linear mixtures of variables were available in the input set because of which they were inadaptable at modeling nonlinear compoundrelations within clinical sectors.

3. RESEARCH METHODOLOGY

In body, Heart performs pumping of blood flowing through the blood vessels existing in circulatory system. Hence, smooth functioning of this organ is very important for the healthy lifestyle. Any type of disorder in this body organ may cause disturbance in other body organs as well. In order to predict heart diseases, researchers generally make use of computer aided data retrieved from huge databases. The use of data mining methodologies and techniques is quite common by multiple businesses. Data mining methods are quite useful for generating predictions about a number of disorders. There are many risk factors that may lead to heart related disorders.

Following are the various phases of heart disease prediction:

A. Data Acquisition: The data is collected from various clinical organizations to perform experiments.

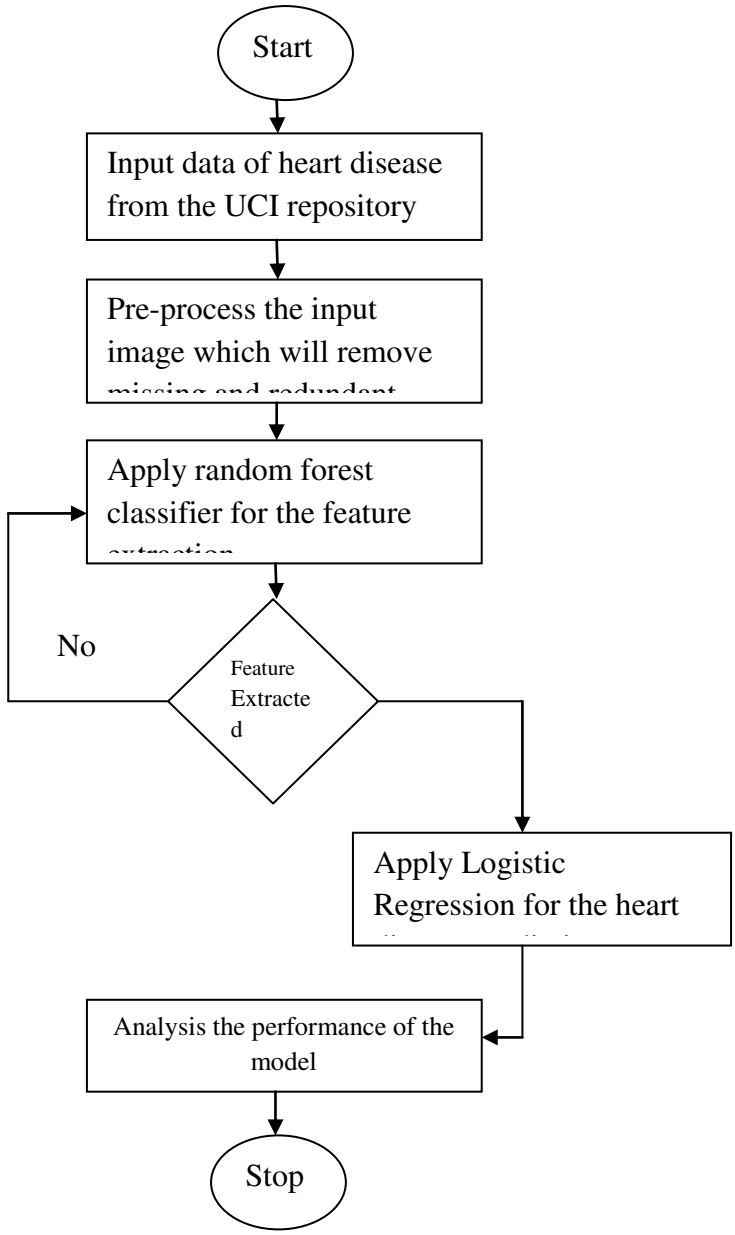
B. Data preprocessing: For applying machine learning techniques such that completeness can be introduced and a meaningful analysis can be achieved on the data, the data preprocessing is performed. This step delivers clean and denoised data for the feature selection process by removing redundant attributes from the dataset for enhancing the efficiency of the training model.

C. Feature selection: This step makes use of a subset comprising extremely unique features for diagnosing heart diseases. These selective features relate to existing class of features. The random forest model is applied for the feature selection in novel scheme.The random forest model takes 100 as the estimator value and generates tree structure of the most relevant features. RF classifier chooses those features which appear most appropriate or significant for predicting heart related disorders.

D. Classification: The mapping of chosen features is carried out to the training model for classifying provided features to make the prediction of disorder possible. A. Here, a kind of heart disease is represented by each separate class. The decision tree model is applied for the classification. The decision takes input of the extracted features. In the research work, two classes are defined which are heart disease and no heart disease. It means that which persons have

probability of heart disease and which don't have probability of heart disease.

Flowchart



Proposed Algorithm

1. Begin with n clusters, each containing one object and we will number the clusters 1 through n.
2. Compute the between-cluster distance D(r, s) as the between-object distance of the two objects in r and s respectively, r, s =1, 2, ..., n. Let the square matrix D = (D(r, s)). If the objects are represented by quantitative vectors we can use Euclidean distance.
3. while (termination criteria is reached) do

   parent selection

   crossover with probability pc

   mutation with probability pm

   decode and fitness calculation

   survivor selection

   find best

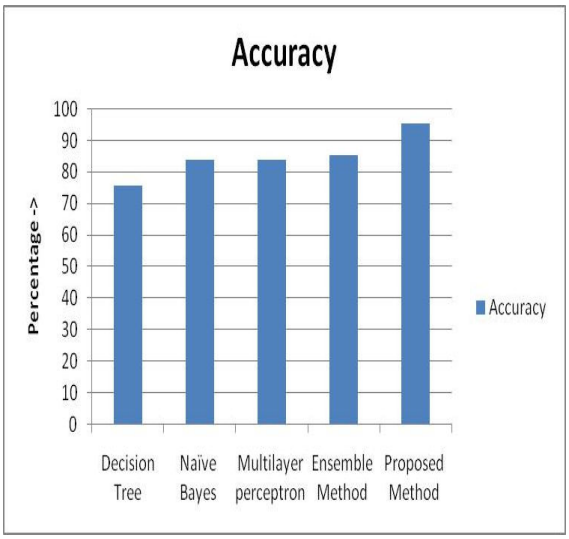
   return cluster

4. Next, find the most similar pair of clusters r and s, such that the distance, D(r, s), is minimum among all the pairwise distances.

5. Merge r and s to a new cluster t and compute the between-cluster distance D(t, k) for any existing cluster k ≠ r, s . Once the distances are obtained, delete the rows and columns corresponding to the old cluster r and s in the D matrix, because r and s do not exist anymore. Then add a new row and column in D corresponding to cluster t.
6. Repeat Step 3 a total of n − 1 times until there is only one cluster left.

4. RESULT AND DISCUSSION

This work uses Cleveland dataset which is most commonly used for predicting heart diseases. This dataset has 14 attributes. In this research work, the implementation and comparison of several models is performed for predicting the heart disease. The DT, Multilayer perception, NB, Ensemble classification method in which random forest, naïve bayes models are combined, proposed models are compared with regard to certain performance parameters.



A variety of models including DT, NB, multilayer perceptron, ensemble and proposed models are compared concerning accuracy. The analytic results reveal that the proposed model achives highest accurcay rate of almost 94.44% by performing better than other classifiers for predicting heart disorders.

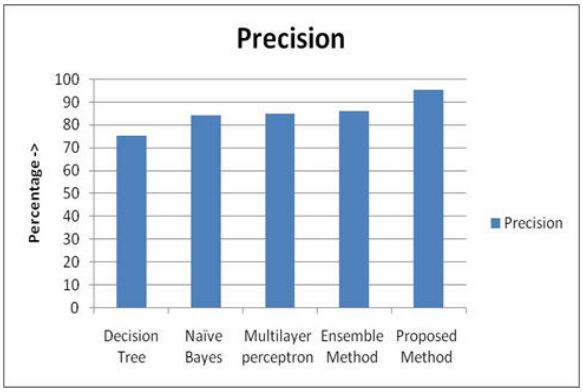


Figure 9: Precision analysis

As shown in figure 9, the various models of including DT, NB, MLP, ensemble and hybrid models are compared in terms of precision. The analytic results reveal that the proposed model achives highest precison rate of almost 95% by performing better than other classifiers for predicting heart disorders.

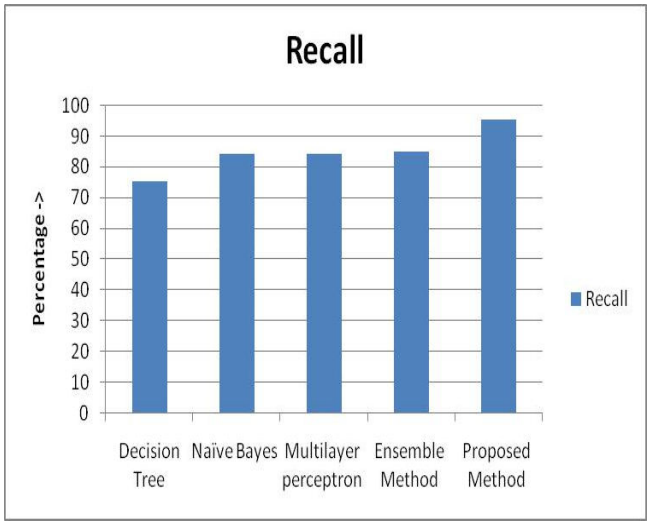


Figure 10: Recall Analysis

As shown in figure 10, the various models like DT, NB, multilayer perceptron, ensemble are compared with the new model in terms of recall. It is analyzed that recall of proposed model for heart disease prediction is approx 95 percent which is higher than the other models.

5. CONCLUSION

The heart disease consists of numerous kinds of diseases due to which various parts of the organ are infected.To conclude, it is analyzed in this work that heart disease prediction is very challenging as the large number of features included in it. The various models are tested for the heart disease prediction like decision tree, naïve bayes, multilayer perceptron, ensemble classifier. The novel model in which the random forest and DTs are integrated is introduced to predict heart disorders. The extraction of features is generated using RF and the decision tree is carried out to carry out classification. The accuracy obtained from the proposed model is computed as 94.44 per cent. The recall and precision obtained from the devised model is computed as 95 percent.

REFERENCES

[1] K.Gomathi, Dr.Shanmugapriyaa, “Heart Disease Prediction Using Data Mining Classification”, 2016, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 4 Issue II

[2] H. Benjamin Fredrick David and S. Antony Belcy, “Heart Disease Prediction using Data Mining Techniques”, 2018, ICTACT Journal on Soft Computing, Volume: 09, Issue: 01

[3] J.Vijayashree and N.Ch. Sriman Narayana Iyengar, “Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques: A Review”, 2016, International Journal of Bio-Science and Bio-Technology, Vol.8, No.4, pp. 139-148

[4] D. Ratnam, P. HimaBindu, V. MallikSai, S. P. Rama Devi and P. Raghavendra Rao, “Computer-Based Clinical Decision Support System for Prediction of Heart Diseases Using Naïve Bayes Algorithm”, 2014, International Journal of Computer Science and Information Technologies, vol. 5, no. 2, pp.2384- 2388

[5] T. Santhanam and E. P. Ephzibah, “Heart Disease Prediction Using Hybrid Genetic Fuzzy Model”, 2015,

Indian Journal of Science and Technology, vol.8, no. 9, pp.797–803

[6] G. Purusothaman and P. Krishnakumari, "A Survey of Data Mining Techniques on Risk Prediction: Heart Disease", 2015, Indian Journal of Science and Technology, vol. 8, no. 12

[7] K. Srinivas, G. RaghavendraRao and A. Govardhan, "Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques", 2010, Proceedings of 5th International Conference on Computer Science & Education, China, pp. 24–27

[6] K. Thenmozhi and P. Deepika, "Heart Disease Prediction Using Classification with Different Decision Tree Techniques", 2014, International Journal of Engineering Research and General Science, vol. 2, no. 6, pp.6-11

[7] S. Chaitrali, Dangare and S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", 2012, International Journal of Computer Applications, vol. 47, no. 10, pp. 0975 –888

[8] J Peter and K. Somasundaram, "An Empirical Study on Prediction of Heart Disease Using Classification Data Mining Techniques", 2012, Proceedings of IEEE International Conference on Advances in Engineering, Science and Management (ICAESM), pp. 514-518.

[9] H. D. Masethe and M. A. Masethe, "Prediction of Heart Disease using Classification Algorithms", 2014, Proceedings of the World Congress on Engineering and Computer Science (WCECS), San Francisco, USA

[10] Jyoti Soni, Ujma Ansari, Dipesh Sharma, Sunita Soni, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction", 2011, International Journal of Computer Applications (0975 – 8887) Volume 17, No.8

[11] Asha Rajkumar, G. Sophia Reena, "Diagnosis of Heart Disease Using Datamining Algorithm", 2010, Global Journal of Computer Science and Technology, Vol. 10 Issue 10

[12] Sunita Soni, O.P.Vyas, "Using Associative Classifiers for Predictive Analysis in Health Care Data Mining", 2010, International Journal of Computer Application, Volume 4, No.5, pages 33-34

[13] Syed Immamul Ansarullah1, Pradeep Kumar Sharma2, Abdul Wahid3, Mudasir M Kirmani, "Heart Disease Prediction System using Data Mining Techniques: A study", 2016, International Research Journal of Engineering and Technology (IRJET), Volume: 03 Issue: 08

[14] Nidhi Bhatla, Kiran Jyoti, "An Analysis of Heart Disease Prediction using Different Data Mining Techniques", 2012, International Journal of Engineering and Technology, Vol.1, issue 8

[15] V. Manikandan and S. Latha, "Predicting the Analysis of Heart Disease Symptoms Using Medical

Data Mining Methods", 2013, International Journal of Advanced Computer Theory and Engineering, Vol. 2, Issue. 2

[16] Anjan Nikhil Repaka, Sai Deepak Ravikanti, Ramya G Franklin, "Design and Implementing Heart Disease Prediction Using Naives Bayesian", 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)

[17] Ankita Dewan, Meghna Sharma, "Prediction of heart disease using a hybrid technique in data mining classification", 2015, 2nd International Conference on Computing for Sustainable Global Development (INDIACom)

[18] Monika Gandhi, Shailendra Narayan Singh, "Predictions in heart disease using techniques of data mining", 2015, International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)

[19] Rashmi G Saboji, "A scalable solution for heart disease prediction using classification mining technique", 2017, International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)

[20] T. John Peter, K. Somasundaram, "An empirical study on prediction of heart disease using classification data mining techniques", 2012, IEEE-International Conference on Advances in Engineering, Science and Management (ICAESM -2012)

[21] Cincy Raju, E Philipsy, Siji Chacko, L Padma Suresh, S Deepa Rajan, "A Survey on Predicting Heart Disease using Data Mining Techniques", 2018, Conference on Emerging Devices and Smart Systems (ICEDSS)

[22] Aakash Chauhan, Aditya Jain, Purushottam Sharma, Vikas Deep, "Heart Disease Prediction using Evolutionary Rule Learning", 2018, 4th International Conference on Computational Intelligence & Communication Technology (CICT)

[23] M. A. Jabbar, ShirinaSamreen, "Heart disease prediction system based on hidden naïve bayes classifier", 2016, International Conference on Circuits, Controls, Communications and Computing (I4C)

[24] Imran Mirza, Arnav Mahapatra, Daryl Rego, Kenneth Mascarenhas, "Human Heart Disease Prediction Using Data Mining Techniques", 2019, International Conference on Advances in Computing, Communication and Control (ICAC3)