Monitoringand Detecting Disease in Human Adults Using Fuzzy Decision Tree and Random Forest Algorithm

R. Dhanalakshimi, C. Geetha, T. Sethukarasi

Abstract: The traditional healthcare involves clinical diagnosis using doctor's expertise and knowledge. It is a challenge to provide proper healthcare in rural and remote areas since they are more likely to travel a long distance to access specialist diagnosis. The number of medical practitioners and facilities are low in these areas making it difficult to provide an expert diagnosis in a significant time interval. The problem can be solved by delivering expert systems to diagnose disease which is built using data mining method and fuzzy logic. The decision trees are widely used in machine learning to predict results. These medical data and expert decision are best represented as the fuzzy data set. The fuzzy decision trees treat fuzzy data and produce simple decision trees. In this project, we built an expert system that diagnoses disease using the random forest algorithm. The fuzzy decision trees are used to increase the accuracy of the diagnosis system. Thus we use Hybrid Fuzzy Decision tree in Random forest algorithm to identify the disease by analyzing the medical records of the patient in this paper.

Keywords: Random Forest, Fuzzy Decision Trees, Health Care, Diagnosis System.

I. INTRODUCTION

In health science, diagnosing critical diseases requires decisions of expert physicians in the field. People living in remote areas cannot access these experts immediately. If certain conditions are not diagnosed and treated correctly in early stages, they become life-threatening. As our world is progressing towards using artificial intelligence and machine learning, designing an expert system that diagnoses the patient's disease has a considerable impact on the health service [1]. The expert system is trained by passing a set of historic patient information records to the machine learning algorithm to diagnose the patient diseases. There are several techniques to design an expert system that diagnoses diseases, here we are using Hybrid fuzzy decision tree in the random forest (HFDTRF) algorithm to solve the problem. The fuzzy trees are used in decision making because a standard decision tree makes a firm decision from crisp dataset while medical data are usually fuzzy [2]-[4]. Thus Fuzzy decision trees can handle fuzziness which is very joint in real-world information [6]. The study of using fuzzy decision trees in the random forest. To classify some medical datasets to simplify an intelligent system to diagnose the disease by the machine learning algorithm such that the result will later be verified by the consultants if preferred. There are many machine learning algorithms which are generally used for sorting problematic [8]-[10]. By providing comprehensible decision-making rules within the medical information record set, decision trees are those that manage to administer satisfactory results. The initial decision tree algorithmic takes only discrete information as input wherever as in the actual web some constant qualities exist in most of the datasets [11].

Thus, some enhancements occur to a transaction with infinite data sets. Still, decision trees had problems in managing in buzzing information. That is, if there is a lesser buzzy in knowledge, that slightly modification attribute standards [3]-[5]. The fuzzy decision tree induction algorithm is introduced to overcome the above issue in applications [13]-[14]. The fuzzy pattern travels on multiple paths is defined by cut points of the given problem which are in-turn defined by fuzzy membership functions and the overlap between them [6]-[7]. The extract information within the form of if-then kind of rules using fuzzy decision tree [12].These classification guidelines are easy to recognize as they are within the manner of human thinking and can be easily represented.

An essential task in any diagnostic system is that the method of trying to spot and determine a possible illness and accomplishment a decision by this method [14]-[16].To deal with misplaced information and through fuzzy details. Must also have the transparency of diagnostic information and also the capacity to describe results. Persons currently are producing a lot of information every single day, so there is a requirement to develop such a classifier which may be accustomed to classify that recently generated information accurately and efficiently[17]-[19]. This expert System primary concentrations on using hybrid fuzzy decision trees that may deal fuzzy information incorporated in Random forest algorithm to provide high-performance results [20]-[22].

In recent years, when the neural network has become increasingly popular in classification problems, because of its relative simplicity of application and its skills to deliver efficient reactions [12]-[14]. In these cases, first decision tree algorithms are succeeded to convey any acceptable consequences. An associate Systems constructed on this method to perform mainly well on any symbolic domains like designation. However, this methodology is not actually applicable [15]-[19]. Consequent developments of this decision tree algorithms could able to deal with numerical information, by selecting a split opinion founded on the educational examples.

Manuscript received January 25, 2019.

R. Dhanalakshimi, Assistant Professor, Department of CSE, R.M.K. Engineering College, Gummidipoondi, Tamilnadu, India.

C. Geetha, Associate Professor, Department of CSE, R.M.K. Engineering College Chennai, Tamilnadu, India.

T. Sethukarasi, Professor, Department of CSE, R.M.K. Engineering College, Gummidipoondi, Tamilnadu, India.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

However in appropriately, this might decrease the accuracy of the classification, particularly for values that are near the bounds of the intervals determined through the training step. To overwhelm the restrictions of decision trees, different approaches that may complement the issues are used [20]-[23]. One amongst these methods is provided by symbolic logic. The fuzzy rule-based structures, the fuzzy instructions are used to deliver simple of accepting, and it also can transfer high-level information. Whereas the fuzzy groups are combined with fuzzy logic and approximate reasoning ways can give the to model fine information details within the problem [22]-[24].

II. LITERATURE SURVEY

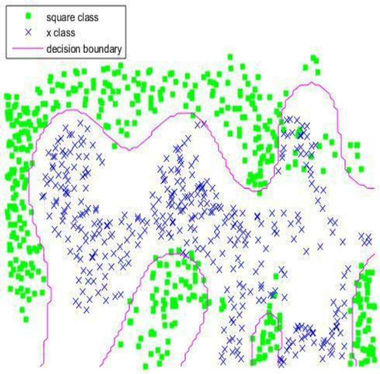
Machine learning algorithms use the classifier system which is highly used to help to solve health care problems. These are used to assist doctors in identifying and forecasting diseases in very initial stages [20]. But, as the medical information set is unorganized, heterogeneous with high dimensions and noise, it is challenging to extract data from medical information records. The idea is to choose an appropriate technique after analyzing all the available methods. There is a rise in the dependence of using medical data for diagnosing a disease. Machine learning algorithms provide great assistance in the medical diagnosis, as interpreting current medical data is very complex[22]-[23]. Machine learning methods may be used for finding big scale multipart biological information analysis as these systems are very active and low-cost thus can be used in solving bioinformatics problems[18]-[19].

A. Support vector Machine technique

One of the general approaches of the statistical education system is SVM, and they are used for classification of medical data Widely. The optimal boundary in the SVM technique is additionally called hyperplane consists of two groups of training vectors which are acquired individually on the probabilistic distribution in a path space. The hyperplane helps to find the boundary that is most distant from the paths nearby the limit in each set. Supporting tracks are the route that is placed close to the hyperplane. There may well be no separating hyperplane if space is not linearly divisible. The kernel purpose is used to solve the problematic. The kernel operates considers the connection between the information.

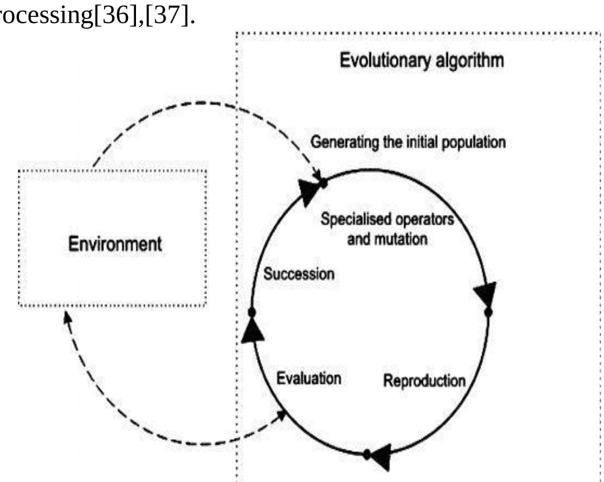
SVM algorithm is an actual numerical learning technique that is planned for the organization because of their high generality performance. Using the essential possible fractions of points of the same category on the identical plane with sets of points from one in all either two classes, SVM can intuitively find the hyperplane. The risk of misclassifying examples of a check the information set is reduced by using optimal separating hyperplane (OSH). Wisconsin Breast Cancer diagnosis (WBCD) used SVM in an exceedingly of investigation on medical diagnosis of breast cancer. They mostly reported high classification accuracies [25]–[30]. Least square SVM was used by Polat and Gunes [26], to produce an efficiency of 99.54%.

Further, for carcinoma designation, network search and have choice SVM model with was proposed [27], [28]. How the best kernel parameters are set and how to selectoptimum input feature for SVM are the two problems confronted using SVM. To worthy analytical and fewer computationally intensive model [29]-[33]. To improve SVM classification, proper model parameters are added to the feature selection.



B. Evolutionary algorithms

Optimal solutions in large and complex spaces are found using an evolutionary and stochastic technique called a Genetic Algorithm (GA). The Genetic Algorithm is motivated by natural evolution: The populace of encrypted applicant solutions is changed through generation's exploitation genetic-like actions like boundary and mutation. The offspring for next generation is created using probabilistic filters at each generation. The preliminary populace is randomly produced to gain a fitness score. So at each generation, an objective function evaluates every candidate solution. To interpret the mined data optimally, the knowledge representation is set by engineers by using evolutionary techniques (presence or absence of disease). The important boundaries are optimally feeding to ANFIS for brain tumor region. The novel neuron-genetic algorithm is used to develop a system which analyzes digital mammograms [34]. The genetic algorithm inputs the significant features which are extracted by the method to a man-made network. This scheme has reached a very fair management [35]-[42]. Thalassemia is diagnosed by ANN+GP [38], chest pain diagnosis by decision tree +GP and abnormalities in the lung are detected using GP+ image processing[36],[37].



Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

C. Swarm Intelligence Technique

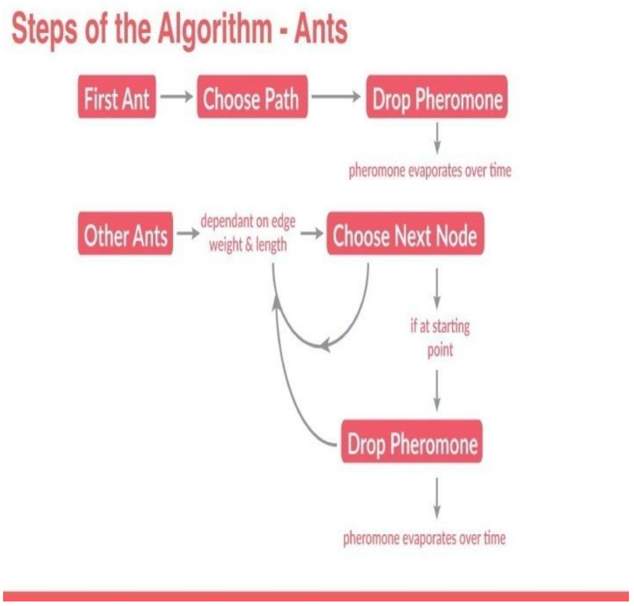
Swarm intelligence (SI) could be a procedure intelligence method to resolve composite actual-global issues. The collective behaviour of persons in a populace interacting nearby with each other and in a reorganized control system With their environment is studied.

The SI is inspired by nature, particularly biological structures.

The managers in SI follows precise, simple rules to imitate natural system. Though any centralized control structure does not regulate the individual agent's behavior.

Thus the local network to an assured mark allows unexpected connections between the mangers that result in an "intelligent" world behavior unknown to the separate managers. Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO) and Artificial Bee Colony (ABC) are some of the popular SI algorithms. The feature selection for medical diagnosis is optimized using Swarm intelligence (SI) algorithms like Particle Swarm Optimization and Ant Colony Optimization.

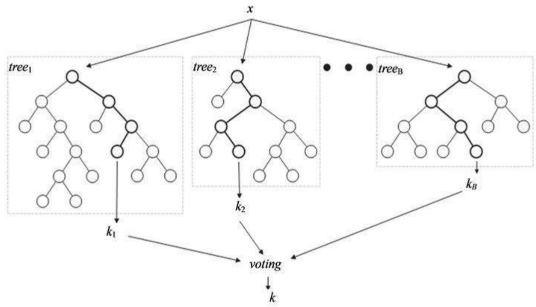
This can will increase the ordering correctness and possesses the development properties required to at least. The employed to the implement gene selection. Gene expression in information ordering problems [46]-[48].



D. Random forest technique

Random forestry algorithm rule is one of the every of the simplest among classification algorithms in use. This algorithmic rule works by creating many decision trees at the coaching period, and it computes the result by predicting the category of the given types of separate trees. Every tree in the random forest algorithm on the values of an arbitrary path which are tested Individually with the identical amount rule on all trees of the forest by combining the tree predictors.

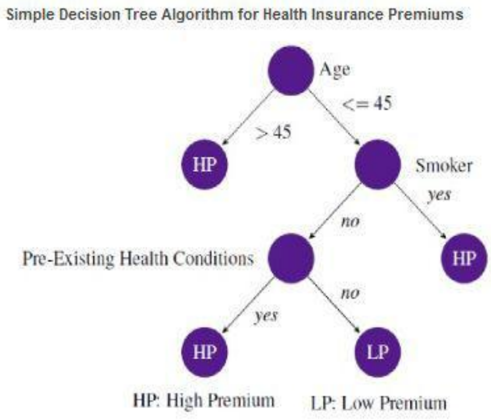
The simple standard of the random forest algorithm is that an assembly of “weak learners” will be combined along to make a "strong learner."



Medical diagnosis is made by researchers using the Random forest as a classifier [53]-[55]. The following four medical data sets are used by the random forest algorithm to obtain optimal results by Ossify [56].

E. Decision tree techniques

The important sorting algorithm is a decision tree. A tree is constructed by a divide and conquer strategy which is pragmatic by the learning algorithm. Each set of attributes gives the sets of instances associated with it. An instance which satisfies the given condition of the problem is represented as a class of value by the leaf node of the decision tree which also contains nodes that represent test performed on the attribute values. The outcome can be either ‘true' or ‘false.' To find and make the assumption of the final class value at the leaf, the nodes are manipulated from root node of the tree to leaf node using certain predefined rules which are derived from the given path. To weed out unnecessary preconditions and duplications in the results, the tree is pruned. Recently a lot of resources are invested in diagnosing disease and using decision on medical classification. Due to its simplicity and interpretability, decision trees are popular for data classification. [65]- [67].The breast cancer diagnosis has been done with using decision trees as a classifier. The prediction of breast cancer is made using three types of classifier proposed by Azar and El-Metwally [62].



The three categories are used particular decision tree SDT, boosted decision tree (BDT) and decision tree forest. They found is a boosted decision tree made improved (98.83%) [62]- [65] Than single decision tree (97.07%)

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The Cerebrovascular disease is predicted with the accuracy of 99.59% by Yeh et al. [60] using a decision tree model which is considerably better than back propagation neural network and Bayesian classifier.

III. PROPOSED SOLUTION

In the Hybrid fuzzy decision tree in a random forest (HFDTRF) algorithm, the tree is created dynamically with the suitable online method. By a substantial modification of random forest algorithm where the decision trees are replaced with a fuzzy decision tree, HFDTRF is constructed. Every tree of HFDTRF is grownup will be explained as follows: allow us to study teaching information size containing M number of records, where M>>m records are bootstrapped from P number of attributes which are tested at random with interchange from the unique data. This model are used for the coaching set for rising the fuzzy decision tree. The most effective splitting on these m attributes is employed to splitting the node and if there are M input variables, a number m<< M is selected such that at each node, m variables are chosen at random out of M. The value of m attributes is held continual through HFDTRF forest developing.

The fuzzy decision tree is developed to the most significant range potential by the algorithm. By sampling interchange member from the training set, a tree forms "in bag" dataset.

It is tested whether the given example information and result are appropriately categorized or not by using the out of bag mistake. It is done with the assistance of out of bag information which usually forms one-third of the "in bag" information.

HFDTRF Algorithm

Input: Dataset

Output: Diagnosed disease data

Step 1: Set Sum of records =m, Sum of attributes =p

Step 2: Let 'm' conclude the quantity of attributes at a node of a decision tree (p<P)

Step 3: For every decision tree do

Select randomly: The subgroup of coaching information that characterizes the M records and uses the rest of information to measuring mistake of the tree, Threshold β for fuzzy decision tree

The fuzzy decision is generated using the below procedure:

While there exist candidate nodes DO

choose one amongst them employing a search scheme , make it's sub node,

Sub-nodes get-together the leaf threshold are leveled as leaf-nodes, then the remaining sub-nodes are viewed as New candidate nodes and therefore the process is continued until the ending measure is met.

End

Before coaching the initial information, the α cut is typically employed for the preliminary information to cut back the fuzziness.

Step 4: For every node of this tree do

Select every which way options to see the choice at this node.

Step 5: End for

Step 6: End for

Fuzzy Knowledge Representation

The aim is to capture the physician's knowledge and experience which is stored as fuzzy data sets. Fuzzy interpretation is employed to improve an algorithmic program that may automatically discovery whether a patient has a specific disease by using fuzzy decision trees from a set of sample diseases. The defuzzified crisp output is used to diagnose illness.

We study a set of pillnesses C and describe a cooperative set of m attribute G related to these illnesses. Usually, we have m>>p. Let:

$$\boldsymbol{C}=\{c_1,c_2,c_3\ldots c_m\}$$

$$\boldsymbol{G}=\{g_1,g_2,g_3\ldots,g_n\}$$

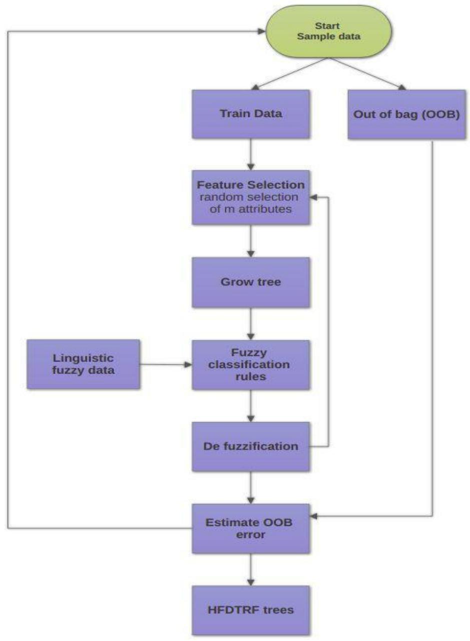
The fuzzy values are chosen from the set:

{Very Low, Low, Moderate, High, Very High}

The set of patient's indications will be obtained as follows:

$$\boldsymbol{S}=\{<g_1,u_1>,<g_2,u_2>,<g_3,u_3>,\ldots,<g_n,u_n>\}$$

Where: u_i is the fuzzy value allocated to the feature gi when Read-through the patient, i=1, …, m.



Shannon Entropy H(S) is used to the quantity of the amount of randomness or uncertainty in given finite set S.

$$H(S)=\sum_{x\in X}p(x)\log_2\frac{1}{p(x)}$$

It provides information regarding the predictability of a particular event. The minor value of Entropy means less uncertainty whereas higher costs mean that high uncertainty. The effective modify in entropy when deciding a specific attribute A for a collection s is given by Data Gain. It is additionally referred to as Kullback-Leibler divergence and denoted by immune IG(S, A).

$$IG(S,A)=H(S)-\sum_{i=0}^{n}P(x)*H(x)$$

The relative modification in entropy concerning the independent variables is calculated using DataGain. Where IG(S, A) is the information gain by applying feature A.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The entire set's entropy is calculated using Shannon Entropy H(S), while the Data Gain measures the Entropy of the set after applying the feature A in it for which probability of event x is P(x).

Fuzzy set operations perform an evaluation of rules. The processes used for OR and AND are t-conorm and q-norm severally.

Q-norm is assumed as,

Q (r, s) =min (r, s) or max (0, r+s-1)

Q-conorm is assumed as

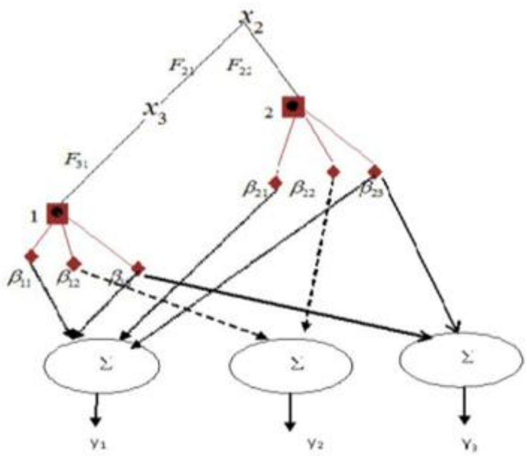
D (r, s) =max (r, s) or min (1, r+s)

Finally, defuzzification is through to supply exact output consequences.

Defuzzification: The scheme needs crisp worth for the management action. So the fuzzy output must be précised to get such worth.

Let us study a fuzzy set,

$$C(y) = \nabla_{1\leq i\leq k} \; [R_{1i}\,(x_i).....R_{ni}\,(x_n) S_i(y)]$$



To obtain a single crisp value for output, the requisite to transmute the membership operate D(y) into a real number C (D). This alteration is termed the defuzzification method.

There are varied defuzzification strategies, the most generally used one is termed Centre of mass defuzzification

Hence the above technique is used to supply products for the method specifically,

$$X^*(y1\ldots\ yn) = C\ (D)$$

Defuzzify the fuzzy set to supply crisp worth. This is referred to as "combine-then-defuzzify". The alpha constant is employed to make your mind the fuzziness of the divisions within the nodes of the fuzzy tree. The 'W' provides the mass allocated to every fuzzy input, while delta provides actual worth.

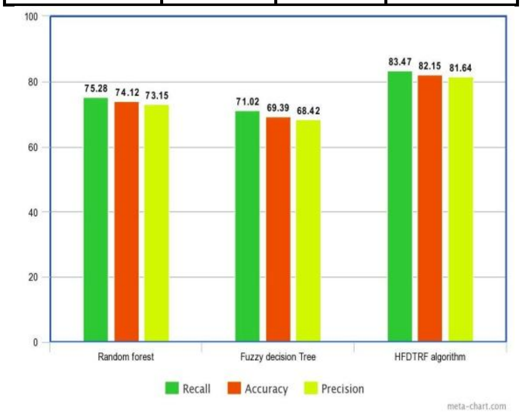
$$\sigma_i = (\sum_{j=1}^{j=k_i} w_{ij}\delta_{ij}) / (\sum_{j=1}^{j=k_i} w_{ij})$$

IV. PERFORMANCE ANALYSIS

This section assesses the show of planned algorithm and production of the devised algorithm is analyzed with Hybrid fuzzy decision tree in a random forest (HFDTRF) algorithmic on the constraints of Precision, and Correctness. Accuracy is that the magnitude relation of the quantity of related records recovered to the whole quantity of related and unrelated documents regained within the database. It is typically communicated as a ratio. Precision is that the percentage of all related and unrelated structures.

The recall is the ration irrelevant structures besides all real geographies. F1 is a harmonic average of accuracy. Association between varied constraints among existing and planned technique. The original random forest produces 74.12% accuracy. While the fuzzy decision trees produce an efficiency of 69.39 %. Finally Hybrid fuzzy decision tree in a random forest (HFDTRF) Produces an accuracy of 82.15 % which is higher than the other two algorithms.

Comparison	Random Forest	Fuzzy Decision Tree	HFDTRF Algorithm
Recall	75.28	71.02	83.47
Accuracy	74.12	69.39	82.15
Precision	73.15	68.42	81.64



V. CONCLUSION

In the healthcare environment, the integration of computer-based systems can be highly benefited by the successful Implementation of the Hybrid fuzzy decision tree in the random forest (HFDTRF) algorithm in medical diagnosis. HFDTRF will assist physicians to diagnose diseases at primary stages, and this will be very useful in developing country like India where the availability of doctor is numbered very minimum, and the mortality rate is pretty high. Technology cannot replace a doctor's expertise and experience, it just saves physician's time by taking care of relatively straightforward yet time-consuming diagnostic tasks, so that doctors can concentrate on the clinically additional demanding process. The use of medical information for diagnosing an illness a sickness is on the growth. Then understanding rare disease is profoundly hard and sometimes diseases are wrongly diagnosed by interns or less experienced physician. This tool can help to dependable assess medical data and thus improve their analytic accurateness, and specificity.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

REFERENCES

1. V. Podgorelec, P. Kokol, B. Stiglic, I. Rozman, "Decision trees: an overview and their use in medicine," Journal of Medical Systems, vol. 26, pp. 445-463, Oct. 2002.
2. L. Rokach, O. Maimon, Data Mining with Decision Trees, World Scientific, 2007.
3. E.I. Papageorgiou, N.I. Papandrianos, D. Apostolopoulos1, P. Vassilakos, "Complementary use of Fuzzy Decision Trees and Augmented Fuzzy Cognitive Maps for Decision Making in Medical Informatics," in Proc. 2008 Int. Conf. on Biomedical Engineering and Informatics, Sanya, Hainan, China, pp.888-892.
4. M. U. Khan, J. P. Choi, H. Shin, M. Kim, "Predicting Breast Cancer Survivability Using Fuzzy Decision Trees for Personalized Healthcare," in Proc. 30th Annual Int. IEEE Conference. on Engineering in Medicine and Biology Society, Vancouver, Canada, 2008, p.5148-5151
5. V. Levashenko, E. Zaitseva, S. Puuronen, "Fuzzy Classified Based on Fuzzy Decision Tree," in Proc. 2007 IEEE Int. Conf. on Computer as
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. Information Technology Solution for Healthcare, K. Zielinski M. Duplaga, D. Ingram Ed. London: Springer, 2006.
8. S. N. Ghazavi, T. W. Liao, "Medical data mining by fuzzy modeling with selected features," Artificial Intelligence in Medicine, vol. 43, pp. 195–206, 2008.
9. Information Technology Solution for Healthcare, K. Zielinski M. Duplaga, D. Ingram Ed. London: Springer, 2006
10. S.Hettich, C.L.Blake, and C.J.Merz, "UCI Repository of ML Databases. Irvine, CA, University of California", Dept. of Information and Computer Science, 1998.
11. C.Z.Janikow, "Fuzzy Decision Trees: Issues and Methods," Journal of IEEE Trans. on Systems, Man, and Cybernetics — Part B: Cybernetics, vol. 28, pp.1-14, Jan. 1998.
12. C. Olaru, L. Whenkel, "A Complete Fuzzy Decision Tree Technique," Fuzzy Sets and Systems, pp.221-254, 2003
13. Y. Yuan, M.J. Shaw, "Induction of Fuzzy Decision Trees," Fuzzy Sets and Systems, vol. 69, pp.125-139, 1995.
14. Evanthia E. Tripoliti, "Automated Diagnosis of Diseases Based on Classification: Dynamic Determination of the Number of Trees in Random Forests Algorithm", IEEE, 2012
15. I. Kononenko, "Machine learning for medical diagnosis: History, state of the art and perspective," Artif. Intell. Med., vol. 23, no. 1, pp. 89–109, 2001.
16. Sarika Pachange, Bela Joglekar, "Random Forests approach for characterizing Ensemble Classifiers," ISSN 2348-4853, 2014
17. Khaled Fawagreh, Mohamed Medhat Gaber & Eyad Elyan, " Random forests: from early developments too recent advancements", ISSN, 2014
18. P. Deepika, P. Vinothini, "Heart Disease Analysis And Prediction Using Various Classification Models-A survey" , ISSN-2250-1991,Volume:4,Issue:3,Mar2015
19. Jiawai Han, Micheline Kamber, "Data Mining: Concepts and Techniques," 2nd Edition, Elsevier, 2006
20. AbdelhmidSalihMohmed Salih1 and Ajith Abraham, "Novel Ensemble Decision Support and Health Care Monitoring System", Journal of Network and Innovative Computing ISSN 2160-2174, Volume 2, 2014
21. Y. Yuan and M. J. Shaw, Induction of fuzzy decision trees, Fuzzy Sets and Systems 69, 125–139(1995).
22. Yeung DS, Wang XZ, Tsang ECC, Learning weighted fuzzy rules from examples with mixed attributes by fuzzy decision trees. In: Proceedings of the IEEE international conference on SMC, Tokyo, Japan, October 12–15,349–354, (1999).
23. Rajen B. Bhatt, Swathi J. Narayanan, Ilango Paramasivam, Khalid M, Approximating Fuzzy Membership Functions from Clustered Raw Data, Proc. IEEE International Conference on Innovations in Social and Humanitarian Engineering(IEEE-INDICON 2012), pages 487-492, Dec 7-9, Kochi, Kerala (2012)
24. Bhatt, Rajen B., and M. Gopal, FRCT: fuzzy-rough classification trees, Pattern analysis and applications 11(1), 73-88, (2008)
25. H. X. Liu, R. S. Zhang, F. Luan, X. J. Yao, M. C. Liu, Z. D. Hu, and B. T. Fan, "Diagnosing breast cancer based on support vector machines," J. Chem. Inf. Comput. Sci., vol. 43, no. 3, pp. 900–907, Jun. 2003.
26. L. Li, H. Tang, Z. Wu, J. Gong, M. Gruidl, J. Zou, M. Tockman, and R. A. Clark, "Data mining techniques for cancer detection using serum proteomic profiling," Artif. Intell. Med., vol. 32, no. 2, pp. 71–83, Oct. 2004.
27. K. Polat and S. Güneş, "Breast cancer diagnosis using least square support vector machine," Digit. Signal Process., vol. 17, no. 4, pp. 694–701, Jul. 2007.
28. M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis," Expert Syst. Appl., vol. 36, no. 2, Part 2, pp. 3240–3247, Mar. 2009.
29. C.-S. Lo and C.-M. Wang, "Support vector machine for breast MR image classification," Comput. Math. Appl., vol. 64, no. 5, pp. 1153–1162, Sep. 2012.
30. M. Çınar, M. Engin, E. Z. Engin, and Y. Ziya Ateşçi, "Early prostate cancer diagnosis by using artificial neural networks and support vector machines," Expert Syst. Appl., vol. 36, no. 3, Part 2, pp. 6357–6361, Apr. 2009.
31. C.-L. Huang, H.-C. Liao, and M.-C. Chen, "Prediction model building and feature selection with support vector machines in breast cancer diagnosis," Expert Syst. Appl., vol. 34, no. 1, pp. 578–587, Jan. 2008.
32. A. E. Hassanien and T. Kim, "Breast cancer MRI diagnosis approach using support vector machine and pulse coupled neural networks," J. Appl. Log., vol. 10, no. 4, pp. 277–284, Dec. 2012.
33. M. Vatankhah, V. Asadpour, and R. Fazel-Rezai, "Perceptual pain classification using ANFIS adapted RBF kernel support vector machine for therapeutic usage," Appl. Soft Comput., vol. 13, no. 5, pp. 2537– 2546, May 2013.
34. T.-C. Chen and T.-C. Hsu, "A GAs based approach for mining breast cancer pattern," Expert Syst. Appl., vol. 30, no. 4, pp. 674–681, May 2006.
35. W. Froelich, E. I. Papageorgiou, M. Samarinas, and K. Skriapas, "Application of evolutionary fuzzy cognitive maps to the long-term prediction of prostate cancer," Appl. Soft Comput., vol. 12, no. 12, pp. 3810–3817, Dec. 2012.
36. G. Schaefer and T. Nakashima, "Data Mining of Gene Expression Data by Fuzzy and Hybrid Fuzzy Methods," IEEE Trans. Inf. Technol. Biomed., vol. 14, no. 1, pp. 23–29, Jan. 2010.
37. A. Das and M. Bhattacharya, "A Study on Prognosis of Brain Tumors Using Fuzzy Logic and Genetic Algorithm Based Techniques," in International Joint Conference on Bioinformatics, Systems Biology and Intelligent Computing, 2009. IJCBS "09, 2009, pp. 348–351.
38. A. Lahsasna, R. N. Ainon, R. Zainuddin, and A. Bulgiba, "Design of a fuzzy-based decision support system for coronary heart disease diagnosis," J. Med. Syst., vol. 36, no. 5, pp. 3293–3306, Oct. 2001
39. G. Schaefer and T. Nakashima, "Data Mining of Gene Expression Data by Fuzzy and Hybrid Fuzzy Methods," IEEE Trans. Inf. Technol. Biomed., vol. 14, no. 1, pp. 23–29, Jan. 2010.
40. W.-J. Choi and T.-S. Choi, "Genetic programming-based feature transform and classification for the automatic detection of pulmonary nodules on computed tomography images," Inf. Sci., vol. 212, pp. 57–78, Dec. 2012.
41. C. Chiu, K.-H. Hsu, P.-L. Hsu, C.-I. Hsu, P.-C. Lee, W.-K. Chiou, T.-H. Liu, Y.-C. Chuang, and C.-J. Hwang, "Mining Three-Dimensional Anthropometric Body Surface Scanning Data for Hypertension Detection," IEEE Trans. Inf. Technol. Biomed., vol. 11, no. 3, pp. 264–273, May 2007.
42. İ. Güler, H. Polat, and U. Ergün, "Combining Neural Network and Genetic Algorithm for Prediction of Lung Sounds," J. Med. Syst., vol. 29, no. 3, pp. 217–231, Jun. 2005.
43. W. Wongseree, N. Chaiyaratana, K. Vichittumaros, P. Winichagoon, and S. Fucharoen, "Thalassaemia classification by neural networks and genetic programming," Inf. Sci., vol. 177, no. 3, pp. 771–786, Feb. 2007.
44. B. Verma and P. Zhang, "A novel neural-genetic algorithm to find the most significant combination of features in digital mammograms," Appl. Soft Comput., vol. 7, no. 2, pp. 612–625, Mar. 2007.
45. U. Bilge, S. Bozkurt, and S. Durmaz, "Application of data mining techniques for detecting asymptomatic carotid artery stenosis," Comput. Electr. Eng., vol. 39, no. 5, pp. 1499–1505, Jul. 2013.
46. W. Zhao and C. E. Davis, "Swarm intelligence based wavelet coefficient feature selection for mass spectral classification: An application to proteomics data," Anal. Chim. Acta, vol. 651, no. 1, pp. 15–23, Sep. 2009.
47. L.-Y. Chuang, H.-W. Chang, C.-J. Tu, and C.-H. Yang, "Improved binary PSO for feature selection using gene expression data," Comput. Biol. Chem., vol. 32, no. 1, pp. 29–38, Feb. 2008.

48. S. Muthukaruppan and M. J. Er, “A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease,” Expert Syst. Appl., vol. 39, no. 14, pp. 11657–11665, Oct. 2012.
49. H.-L. Chen, B. Yang, G. Wang, S.-J. Wang, J. Liu, and D.-Y. Liu, “Support Vector Machine Based Diagnostic System for Breast Cancer Using Swarm Intelligence,” J. Med. Syst., vol. 36, no. 4, pp. 2505–2519, May 2011.
50. H. J. Escalante, M. Montes-y-Gómez, J. A. González, P. Gómez-Gil, L. Altamirano, C. A. Reyes, C. Reta, and A. Rosales, “Acute leukemia classification by ensemble particle swarm model selection,” Artif. Intell. Med., vol. 55, no. 3, pp. 163–175, Jul. 2012.
51. L.-Y. Chuang, H.-W. Chang, C.-J. Tu, and C.-H. Yang, “Improved binary PSO for feature selection using gene expression data,” Comput. Biol. Chem., vol. 32, no. 1, pp. 29–38, Feb. 2008.
52. S. Muthukaruppan and M. J. Er, “A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease,” Expert Syst. Appl., vol. 39, no. 14, pp. 11657–11665, Oct. 2012.
53. J. Maroco, D. Silva, A. Rodrigues, M. Guerreiro, I. Santana, and A. de Mendonça, “Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests,” BMC Res. Notes, vol. 4, no. 1, pp. 299, Aug. 2011.
54. K. M. Cherry, S. Wang, E. B. Turkbey, and R. M. Summers, "Abdominal lymphadenopathy detection using the random forest," Proc. SPIE 9035, vol. 9035, pp. 90351G-90351G-6, March 2014
55. Bhatt, Rajen B., and M. Gopal, FRCT: fuzzy-rough classification trees, Pattern analysis and applications 11(1), 73-88, (2008).
56. Simon Bernard, Laurent Heutte, Sehastien Adam, "Forest- RK: A New Random Forest Induction Method", ICIC (2), Springer, pp.430-437, Lecture Notes in Computer Science, vol.5227, 2009
57. A. Oct, “Enhanced Cancer Recognition System Based on Random Forests Feature Elimination Algorithm,” J. Med. Syst., vol. 36, no. 4, pp. 2577–2585, May 2011.
58. C. Nguyen, Y. Wang, and H. N. Nguyen, “Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic,” J. Biomed. Sci. Eng., vol. 06, no. 05, pp. 551–560, 2013.
59. P. Padilla, M. López, J. M. Górriz, J. Ramírez, D. Salas-González, I. Álvarez, and Alzheimer‟s Disease Neuroimaging Initiative, “NMF-SVM based CAD tool applied to functional brain images for the diagnosis of Alzheimer‟s disease,” IEEE Trans. Med. Imaging, vol. 31, no. 2, pp. 207–216, Feb. 2012.
60. R. C. Barros, M. P. Basgalupp, A. A. Freitas, and A. C. P. L. F. de Carvalho, “Evolutionary Design of Decision-Tree Algorithms Tailored to Microarray Gene Expression Data Sets,” IEEE Trans. Evol. Comput., vol. 18, no. 6, pp. 873–892, Dec. 2014.
61. D.-Y. Yeh, C.-H. Cheng, and Y.-W. Chen, “A predictive model for cerebrovascular disease using data mining,” Expert Syst. Appl., vol. 38, no. 7, pp. 8970–8977, Jul. 2011.
62. N. Cruz-Ramírez, H. G. Acosta-Mesa, H. Carrillo-Calvet, L. Alonso Nava-Fernández, and R. E. Barrientos-Martínez, “Diagnosis of breast cancer using Bayesian networks: A case study,” Comput. Biol. Med., vol. 37, no. 11, pp. 1553–1564, Nov. 2007.
63. D. Delen, G. Walker, and A. Kadam, “Predicting Breast Cancer Survivability: A Comparison of Three Data Mining Methods,” Artif Intell Med, vol. 34, no. 2, pp. 113–127, Jun. 2005.
64. J. M. Luk, B. Y. Lam, N. P. Y. Lee, D. W. Ho, P. C. Sham, L. Chen, J. Peng, X. Leng, P. J. Day, and S.- T. Fan, “Artificial neural networks and decision tree model analysis of liver cancer proteomes,” Biochem. Biophys. Res. Commun., vol. 361, no. 1, pp. 68–73, Sep. 2007.
65. C.-Y. Fan, P.-C. Chang, J.-J. Lin, and J. C. Hsieh, “A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification,” Appl. Soft Comput., vol. 11, no. 1, pp. 632–644, Jan. 2011.
66. D.-Y. Yeh, C.-H. Cheng, and Y.-W. Chen, “A predictive model for cerebrovascular disease using data mining,” Expert Syst. Appl., vol. 38, no. 7, pp. 8970–8977, Jul. 2011.
67. N. Cruz-Ramírez, H.-G. Acosta-Mesa, H. Carrillo-Calvet, and R.-E. Barrientos-Martínez, “Discovering interobserver variability in the cytodiagnosis of breast cancer using decision trees and Bayesian networks,” Appl. Soft Comput., vol. 9, no. 4, pp. 1331–1342, Sep. 2009.

Dhanalakshmi, Assistant Professor, CSE, R.M.K. Engineering College, Chennai
Ph.D. Student, Anna University, Chennai.
M.E. Anna University, Chennai
Research Interests:
Data Mining Neural Networks and Machine Learning

Dr.C. Geetha, Associate Professor, CSE, R.M.K. Engineering College, Chennai Ph.D. Manonmaniam Sundaranar University M.E. Sathyabama Institute of Science & Technology
Research Interests:
Wireless Sensor Networks, IoT

Dr.T. Sethukarasi, Professor and Head, CSE, R.M.K.Engineering College,
Chennai Ph.D. College of Engineering, Anna University, Chennai
M.E. Anna University, Chennai
Research Interests:
Data Mining Neural Networks and Machine Learning

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication