CLDC: Efficient Classification of Medical Data Using Class Level Disease Convergence Divergence Measure

K.Ananthajothi, M.Subramaniam

Abstract*: The problem of medical data classification is analyzed and the methods of classification are reviewed in various aspects. However, the efficiency of classification algorithms is still under question. With the motivation to leverage the classification performance, a Class Level disease Convergence and Divergence (CLDC) measure based algorithm is presented in this paper. For any dimension of medical data, it convergence or divergence indicates the support for the disease class. Initially, the data set has been preprocessed to remove the noisy data points. Further, the method estimates disease convergence/divergence measure on different dimensions. The convergence measure is computed based on the frequency of dimensional match where the divergence is estimated based on the dimensional match of other classes. Based on the measures a disease support factor is estimated. The value of disease support has been used to classify the data point and improves the classification performance.

Keywords*: High Dimensional Clustering, Classification, Medical Data, Convergence, Divergence, Disease Prediction

I. INTRODUCTION

The recent development of medical solutions and information technology represent the medical data in different dimensions. The size of medical data has no restriction and the volume of the medical data also higher. The most organization stores the medical data in various forms, however, it is necessary to index the medical data with higher accuracy. The performance of classification depends on how the data points are grouped. Grouping similar and related medical data under a single label has been named as clustering. Consider the data set Ds contains X number of data points and clustering the data points of Ds, in to C ={c1,c2,..Cn} of classes where each class Ci contains x number of data points which is a subset of Ds. In general, clustering is performed using may techniques where each uses different measures. The popular K-means algorithm measures the similarity between data points according to Euclidean distance measure to group them. Similarly, the Fuzzy C means algorithms computes the range values and mean value in identifying the class of the data point.

The medical data has large number of dimensions and the application of k-means and Fuzzy clustering would not produce efficient results in clustering. So, it is necessary to consider the strategic measures and approaches in clustering. For any dimension, it will be moving the data point toward a cluster where it will be moving the data point away from the clustering. This corollary makes the clustering to consider the maximum number of dimensions in clustering.

Classification is the process of classifying a given data point Di between N number of Disease classes Dc. In general, the classification is done according to similarity between the data points of any cluster or disease class with the input data point. The similarity between all the disease classes will be estimated for the input data point. At last a disease class has been selected as the result. Fuzzy rule based classifications are available, which maintains set a fuzzy values for all the properties of the data point. The rule has been maintained for all disease class. Based on the fuzzy rule available, the fuzzy algorithm estimates similarity of data point in all the dimensional values. Similarly, the classification is performed using, support vector machines which estimates support value based on input vectors. Number of algorithms available for the classification of medical data. But the efficiency of classification is highly depending on the similarity measure being used. The performance of fuzzy algorithm depends on the range value of each dimension, if the dimensional value of data point is more scattered then the range value will be higher which produces poor classification. So, the selection of dimension and the similarity measure plays the important role in classification. This paper, concentrates on the similarity measure being used for both clustering and classification. Because to use an similarity measure for classification, the clustering should be done based on that. The convergence of any data point towards a disease class represent the attraction of the data point towards the disease class considered. It can be measured based on the number of samples being similar in disease class considered. The samples similarity can be measured based on the values of all the dimension similarity in distance. Based on the distance measure the convergence value is measured. The Divergence measure of any data point towards a disease class is estimated based on the similarity of the same data point towards the data points of other disease class. When a data point has higher convergence measure with other classes than particular class, then it has higher divergence measure from the class considered. By computing the convergence and divergence measures, you can measure the similarity of the data points towards any class.

Revised Manuscript Received on August 09, 2019.
***** Correspondence Author
K.Ananthajothi*, Assistant Professor, Department of Computer Science and Engineering, Misrimal Navajee Munoth Jain Engineering College, Chennai, India. Email:kanandjothime@gmail.com
M.Subramaniam, Professor, S.A. Engineering College, Chennai, India. Email:drms@saec.ac.in

This would improve the performance of classification.

This paper present an efficient classification algorithm based on the two measures discussed above. The abstract method is detailed in the coming sections.

II. Related Works

There exist number of classification approaches specified and this section discusses set of methods related to the problem. In [1], the author discussed the effect of preprocessing while classifying the medical data. The author conclude that the effect of classification is depends on the combination of preprocessing approach being considered. Also it depends on the way you handle the missing values and how you handle the numeric values. The author presented a case study based on ant colony optimization technique.

Different metrics of medical data classification and how it has been adapted for diagnosis of disease and prediction of any disease based on different models is presented in [2]. The application of dispersed medical data in decision making is discussed in [3].The author used both local data and global data with collision. Initially the classification is performed based on the local data using probability vectors of different classes. The same probability vectors has been used to perform clustering and applied a conflict analysis. The problem of un certainty in the number of attributes in medical data classification and diagnosis is presented in [4]. The method uses the neuro fuzzy technique to generate rules by identifying the number of attributes of the data set. Also, the disease prediction is performed based on the fuzzy rules generated. In [5], the author presented a medical content classification algorithm which uses rules. The method preprocesses the data and normalizes them to produce rules. Based on the rules generated, the data classification is performed. An application of Hybridized Smote Technique with Rough set has been presented in [6]. The method uses rough set theory for classification. Also, the method uses maximum distance to handle the imbalanced data. The method uses a assessment metrics for the evaluation of balance of data. For the prediction of heart disease an efficient algorithm has been described in[7]. The method uses the k-means algorithm for the clustering of big data. The k means algorithm in turn uses the ID3 algorithm for clustering. The method produces efficient results on clustering. The application of local difference in classification has been presented in [8]. The method first generates contrast patterns based on which a set of rules has been generated. According to that, a set of contrasting rules has been identified to measure the distance between data points. The contrast pattern based rule and classifier produces efficient results on classification. In [9], the author presented least mean square algorithm based learning approach for classification and disease prediction. The method uses different weights and different scientific approaches in classification. In [10], author claims that the selection of exact combination of attributes has higher impact in the classification of any data set. Also, the numeric attributes should have tackled efficiently to achieve higher classification performance. The author presented a case study towards this corollary.

An classification technique for the problem of heart disease prediction is presented to be worked over medical data in [11]. The method uses the patient history and generates a graph and produces set of words in understandable manner. The method uses naïve bayes and ID3 algorithms for clustering the data set. Also, a decision tree has been adapted to the problem of clustering. For the classification of High speed data in real time, an Nearest Neighbor approach is presented using Spark in [12]. The method performs classification in incremental manner using the NN algorithms. The method produces good results based on the spark. An ensemble based classification algorithm is presented for diabetic prediction in [13].The method uses SMOTE tool for the classification and also adapted the various classification algorithms for the evaluation of disease prediction on medical data set. In [14], an genetic algorithm based classification towards heart disease has been presented. The method applies KNN classifier on the government medical data and performs classification. The application of Disease Prediction on the big data has been presented in [15]. The health care community data has been collected and the method applies the streamline machine learning technique over the data collected. The method predicts the chronic disease based on frequency measures on the community data. The association of gene on occurrence of any disease using text mining is presented in [16]. The method computes term frequency and term weight for each genes and their associated. Based on the weight, the gene being influences the disease has been predicted. The genetic algorithm has been used for heart disease prediction [17]. The method generate recurrent fuzzy neural network for training the network where the genetic algorithm has been used to estimate the weights and prediction. In [18], the author presented a supervised algorithm for the clustering of genes based on their mutual information. The method has been evaluated using different gene data sets and data sets of various diseases. Similarly, in [19], a hybrid algorithm for cancer prediction is presented which combines both SVM and gene sets. In [20], to identify the pathway for cancer, a statistical method is presented which uses probability based principle component analysis with sparse data points named (SPPCA). In [21], a multilevel approach for medical data classification is presented which uses influence measure in different level to identify the class of data points in an iterative manner. All the above discussed methods suffer to produces higher classification efficiency and produces les s disease prediction accuracy.

III. CLDC Based Classification

The proposed class level divergence and convergence measure based classification algorithm, first preprocess the input data set by removing the noisy records. The preprocessed data set is used to cluster the data points using CLDC clustering. For clustering, each data point has been measured with DC measures. Based on the DC measures, a single class has been selected and indexed. Similarity for classification, the same DC measure has been estimated and finally a class level similarity (CLS) has been measured. The class with higher CLS has been selected as the target class.

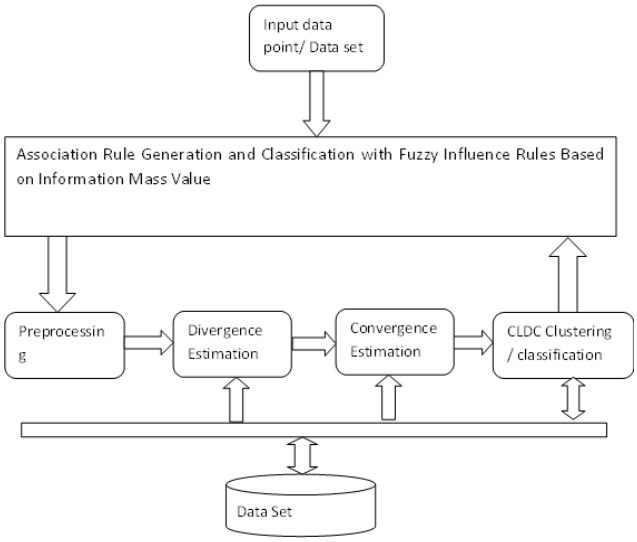


Figure 1: Functional Block Diagram of proposed CLDC based classification

The functional abstract view of CLDC classification algorithm is presented in Figure 1. It shows the various stages involved in the system.

A. Preprocessing

Given a medical data set X, would contain Y number of data points belongs to Z number of categories. Each data point x, would contain N number of values which is same for all the categorical points. Preprocessing is to identify the data point x and verifying the completeness of the data point for all the dimension n. If any of the dimension n of x is identified as incomplete or it have null value, then it is considered as incomplete or noisy which is eliminated. Such noise removed data set X will by used to perform CLDC clustering and CLDC classification.

First the dimension list has been identified from input data set as follows:

Dimension set Ds

$$= \int_{i=1}^{size(X)} \int_{j=1}^{size(Z)} if(Ds \nexists j), (Ds \cup J), null \quad (1)$$

The above equation identifies each dimension from each data point of data set X, if any dimension identified as not contained in DS, then it will be added to the dimension set Ds, otherwise it will be neglected.

Now its time to verify the completeness of the data points of X.

Let's take a data point x, and the completeness will be verified as follows:

$$\int_{i=1}^{size(DS)} if(\ x \in \forall Ds(i), true, false) \quad (2)$$

Equation (2), returns a true value only when the data point x contains all the dimensions from dimension set Ds, otherwise it will return a false value. According to the value returned by (2), the data point will be either kept or removed from data set.

If true then

Leave the point.

Else

$$X = X \cap x \quad (3)$$

End

Equation (3), removes the incomplete data point x from the input data set. The completeness is verified according to the possession of all dimensions or feature and their values.

The noise removed set X has been utilized for clustering and classification by the below algorithms.

B. CLC Estimation

The class level convergence measure represent the attraction of the input data point towards the data points of the target class considered. Consider the data point x, has been counted for measuring CLC measure towards a class C {y1,y2,…,yn}, then the CLC measure is estimated according to the attraction of the data points between x and data points of C. It has been measured based on the distance between the mean and standard deviation values of each dimension of data points. The estimated CLC measure has been used to perform clustering as well as classification.

First, the mean value of each dimension of data points of the class C has been measured as follows:

$$\text{Mean set Ms} = \int_{i=1}^{size(DS)} \frac{\sum_{j=1}^{size(C)} C(j)(i)}{size(C)} \quad (4)$$

Second, the standard deviation of all the values of each dimension of data point has been measured as follows:

$$\text{StdSet Stds} = \int_{i=1}^{size(DS)} Stdev(\sum_{j=1}^{size(C)} C(j)(i)) \quad (5)$$

Algorithm:

Input: Class C, Data Point x

Output: CLC

Start

Read Data points of class C.

Read input data point x.

Mean set Ms = Compute mean set for Class C using (4)

StdSet Stds = Compute standard deviation set for the class C using (5).

Initialize Count =0;

For each dimension n

if Dist(Ms(n),x(n))<stds(n) then

count = count+1.

End

End

Compute convergence measure CLC = Count/N (6)

Stop

The CLC algorithm estimates the class level convergence measure which represent the attraction of data point from the data points of the class C. The estimated measure will be used to perform clustering and classification.

C. CLD Estimation

The class level divergence measure represent the deviation of the data point x according to class C considered. . Consider the data point x, has been counted for measuring CLD measure towards a class C {y1, y2,…, yn}, then the CLD measure is estimated according to the deviation of the data points between x and data points of C. It has been measured based on the distance between the mean and standard deviation values of each dimension of data points. The estimated CLC measure has been used to perform clustering as well as classification.

Algorithm:
Input: Class C, Data Point x
Output: CLD
Start
Read Data points of class C.
Read input data point x.
Mean set Ms = Compute mean set for Class C using equation (4)
StdSet Stds = Compute standard deviation set for the class C using Equation (5).
Initialize Count =0;
For each dimension n
if Dist(Ms(n),x(n))>stds(n) then
count = count+1.
End
End
Compute convergence measure CLD = Count/N
Stop

The CLD algorithm estimates the class level divergence measure which represent the deviation of data point regarding the class C. The estimated measure will be used to perform clustering and classification.

D. CLDC Clustering

The CLDC clustering is meant for grouping given data set X into number of classes C according to the convergence and divergence measures of data points. To perform clustering, first the algorithm initializes number of class and for each of the class c, a set of points are initialized. In the second stage, the method computes CLC and CLD measures for each data point x towards all the classes of C. Based on the estimated CLC and CLD measures, the method compute the CLDC weight for the input point x. Based on computed CLDC weight, a single class has been selected and the data point has been indexed. This will be iterated for number of time for each class of points indexed until there is no movement of point between any classes.

Algorithm:
Input: Data Set X, Cluster set Cs.
Output: Cluster Set Cs.
Start
Read X, Cs.
Initialize number of classes to the Cs.
For each class c
Add random samples from X.
$Cs© = \int Random((X \nexists c), R)$ (7)

Equation (7) will randomly select R number of samples and place then in the class c.

End
For each data point x
For each class c
Compute CLC = CLCEstimation(x, c)
Compute CLD = CLD-Estimation(x, c)
Compute CLDC Weight = CLC×CLD (8)
End
Choose the class c with higher CLDC weight.
Cs©(Max(CLDCW))
$= \sum x(Cs©(Max(CLDCW))) \cup x$ (9)
End
Initialize count=1
While count>0
Count=0
For each class c
For each data point x
Compute CLDC weight with the class c
For each other class Oc
Compute CLDC weight.
End
Class Oc=Choose the class with higher CLDC Weight.
If OC!=c then
Swap x
Count = count +1
End
End
End
End
Stop

The algorithm presented above performs CLDC clustering in iterative manner till there is a swap of data point between any two clusters or classes. Generated cluster will be used to perform classification in the next stage.

E. CLDC Classification

In this stage, for a given input data point x, the method estimates CLC and CLD measures towards different classes of C. Using both the measures, the method estimates CLDW measure. Based on the CLDW measure, the method select a single target class to which the input data point falls. The medical data has been assigned with the selected class.

Algorithm: Input data point x, Cluster Set Cs.
Output: class c
Start
Read x, Cs.
For each class of Cs
Estimate CLC measure
$= \int_{i=1}^{size(Cs)} CLCEstimation(x, Cs)$
Estimate CLD measure
$= \int_{i=1}^{size(Cs)} CLDEstimation(x, Cs)$
Estimate CLDW weight = CLC×CLD
End
Class c = Max (∀Cs(c)) (10)
Stop

The algorithm detailed above estimate the CLDW weight for the given input data point with the classes available. Finally a disease class is selected and the selected class has been tagged with the input point.

IV. RESULT AND DISCUSSION

The proposed CLDC Classification algorithm has been implemented using advanced java. The method has been validated for its efficiency in classification using different data sets. The CLDC classification algorithm leverages the performance of classification and prediction. The evaluation result has been described below with the details of evaluation table 1.

Table 1: Description of Data set used

Key	Value
Name of Data set	UCI-Diabetic, UCI-Hepatitis
Total Samples	23174, 155
Total Features	275,19
Total Classes	5,2

The details of data set used for the evaluation of CLDC algorithm is presented in Table 1. There are totally 275,19 number of samples grouped under five classes where each have 274 features in UCI Diabetic data set. Similarly, the UCI Hepatitis data set has the samples of 155 which have covered by 2 classes and each sample has given with 19 features. The diseases does not share common dimensions and they will be differ at each disease.

The performance of the CLDC medical data classification algorithm is evaluated under different factors. Such results obtained from evaluation have been compared with different baseline methods.

The time complexity is the value of time in seconds consumed by the algorithms in classifying the input sample.

Time Complexity = (Total Time Taken)/(Number of samples)

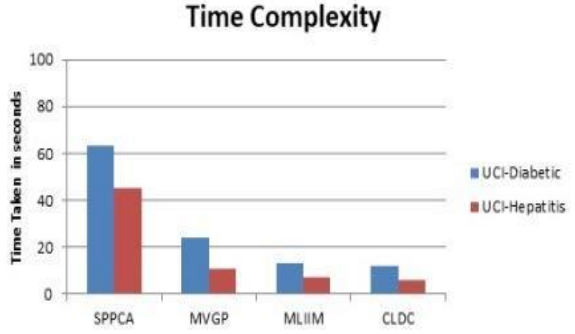
The complexity in time produced by the CLDC algorithm has been measured and presented in Graph 1. By viewing the Graph 1, it can be understood that the proposed CLDC algorithm has achieved less time complexity compare to other approaches.

The ratio in classifying the sample as false class is measured with false ratio and presented. It is measured according to the number of false classification made by the algorithm for a given number of samples.

Table 2: Performance Result on Time Complexity

Time Complexity in seconds		
Method	UCI-Diabetic	UCI-Hepatitis
SPPCA	63	45
MVGP	24	11
MLIIM	13	7
CLDC	12	6

Table 2, present the result of performance analysis on time complexity produced by different methods. The proposed CLDC algorithm has produced less time complexity in medical data classification towards all the data set considered.



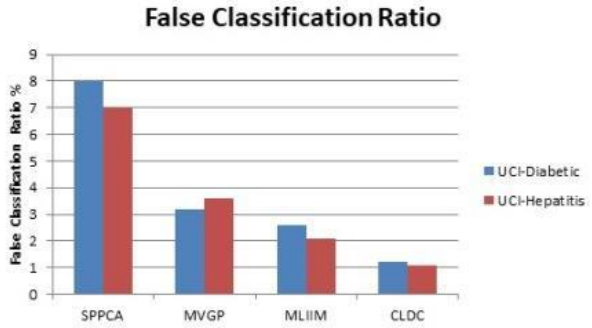
Graph 1: Performance on Time Complexity

False Classification Ratio = (Number of Fase classification)/(Total Number of samples)×100

Table 3: Performance result on False Classification Ratio

False Classification Ratio %		
Method	UCI-Diabetic	UCI-Hepatitis
SPPCA	8	7
MVGP	3.2	3.6
MLIIM	2.6	2.1
CLDC	1.2	1.1

The Table 3, present the result of performance analysis on false classification ratio produced by different methods. The proposed CLDC algorithm has reduced the false classification ratio than other methods.



Graph 2: Performance on False Classification Ratio

The ratio of false classification made by the algorithms has been computed and presented in Graph 2. The CLDC algorithm has reduced the ratio of false classification than the previous MLIIM algorithm.

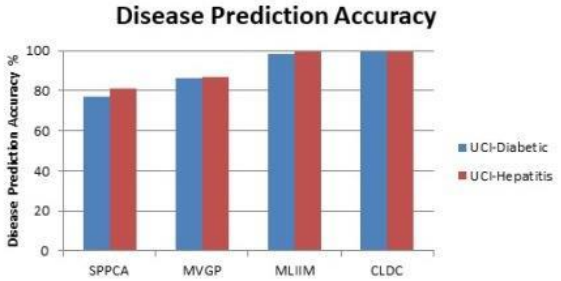
The accuracy of disease prediction represent the efficiency of the algorithm. It has been measured in two ways as classifying true as positive and true as negative.

Disease Prediction Accuracy = (TruePositive+TrueNegative) / (Total No of classification)

Table 4: Performance result on Disease Prediction Accuracy

Disease Prediction Accuracy %		
Method	UCI-Diabetic	UCI-Hepatitis
SPPCA	77	81
MVGP	86	87
MLIIM	98.6	99.6
CLDC	99.3	99.7

The performance on disease prediction accuracy has been measured and presented in Table 4. The proposed CLDC algorithm has produced higher disease prediction accuracy compare to other methods.



Graph 3: Performance on disease prediction accuracy

The accuracy in disease prediction has been measured for different algorithms and presented in Graph 3. The CLDC algorithm has produced higher accuracy in disease prediction compare to other methods.

V. CONCLUSION

This paper presented an class level divergence and convergence measure based medical data classification algorithm. First the method preprocess the data to perform noise removal and the noise removed data point has been clustered into number of classes using Class Level Divergence and Class Level Convergence measures. Grouping the data points of data set according to CLDC measure improves the clustering accuracy. Using the values of CLD and CLC measures, the method compute CLDW for different class of data points. The CLDW measure represent the fitness of data point to the class. Finally a single class to which the data point has more CLDW value has been selected and indexed. The same set of measures has been used to perform disease prediction and classification. The proposed CLDC algorithm has improved the performance of classification up to 99.3 % and the false ratio has been reduces up to 0.7%.

ACKNOWLEDGMENT

This work was done in “Big Data and Cloud Computing Lab” at Dept. of Information Technology, S.A. Engineering college. The authors would like to thank, Department of Science & Technology, Ministry of Science & Technology, Govt. of India, for granting the fund under “Fund for Improvement of S&T Infrastructure in Universities and Higher Educational Institutions (FIST) Program – 2014”, Grant Sanction order vide: SR/FST/COLLEGE-239/2014, dated: 21st Nov 2014, for establishing “Big Data and Cloud Computing Lab” for strengthening the existing institutions S&T infrastructure and support for advancement in research works.

REFERENCES

1. Sarab Almuhaideb,Mohamed El Bachir Menai, Impact of preprocessing on medical data classification, Springer, Frontiers of Computer Science 2016, Volume 10, Issue 6, pp 1082–1102.
2. Sarab Al-Muhaideb, Mohamed El Bachir Menai Hybrid Metaheuristics for Medical Data Classification, Springer, Hybrid Metaheuristics pp 187-217.
3. Małgorzata Przybyła-Kasperek, Alicja Wakulicz-Deja, Global Decisions Taking on the Basis of Dispersed Medical Data, Springer, International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing RSFDGrC 2013
4. Iryna Pliss,Iryna Perova, Diagnostic Neuro-Fuzzy System and Its Learning in Medical Data Mining Tasks in Conditions of Uncertainty about Numbers of Attributes and Diagnoses, Springer, Automatic Control and Computer Sciences November 2017, Volume 51, Issue 6, pp 391–398
5. Agnieszka Mykowiecka Anna Kupść Małgorzata Marciniak, Rule-Based Medical Content Extraction and Classification, Springer, Intelligent Information Processing and Web Mining pp 237-245, 2017.
6. Nadir Mustafa ; Jian-Ping Li, Medical data classification scheme based on hybridized SMOTE technique (HST) and Rough Set technique (RST), IEEE Conference on cloud computing and big data analytics, 2017.
7. Tejaswini U. Mane, Smart heart disease prediction system using Improved K-means and ID3 on big data, IEEE Conference on Data Management, Analytics and Innovation (ICDMAI), 2017.
8. Marharyta Aleksandrova ; Oleg Chertov ; Armelle Brun ; Anne Boyer , Contrast classification rules for mining local differences in medical data, IEEE Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2017.
9. Sara Belaroucia), Mohammed Amine Chikh ,Medical imbalanced data classification, ASTES, Volume 2, Issue 3, Page No 116-124, 2017.
10. SarabAlMuhaideb, Mohamed El BachirMenai, An Individualized Preprocessing for Medical Data Classification, Elsevier, Computer Science, vol.82, 2016.
11. S. Ranganatha ; H. J. Pooja Raj ; C. Anusha ; S. K. Vinay , Medical data mining and analysis for heart disease dataset using classification techniques, IEEE Conference on Research & Technology in the Coming Decades (CRT 2013).
12. Sergio Ramires-gallego, Bartosz Krawczvk, Nearest Neighbor Classification for High-Speed Big Data Streams Using Spark, IEEE Transaction on Systems, Man, Cybernetics, Systems, Vol.47, Issue 10, 2017.
13. Manal Alghamdi, Mouaz Al-Mallah, Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford ExercIse Testing (FIT) project, PLOS, 12(7), 2017.
14. M. Akhiljabbar, B.L.Deekshatulu, PritiChandra, Classification of Heart Disease Using K- Nearest Neighbor and Genetic Algorithm, Elsevier Procedia Technology , vol.10, pp:85-94, 2013.
15. Min Chen ; Yixue Hao ; Kai Hwang ; Lu Wang ; Lin Wang , Disease Prediction by Machine Learning Over Big Data From Healthcare Communities, IEEE Access, vol. 5, pp:8869-8879, 2017.
16. Jie Zhou, Bo-quan Fu ,The research on gene-disease association based on text-mining of PubMed,Springer, BMC Bioinformatics, 2018.
17. Kaan Uyara, Ahmet İlhana, Diagnosis of heart disease using genetic algorithm based trained recurrent fuzzy neural networks, Elsevier, Procedia Computer Science 120 (2017) 588–593.
18. PradiptaMaji and Chandra Das, “Relevant and Significant Supervised Gene Clustersfor Microarray Cancer Classification”, IEEE Transactions on Nanobioscience, vol. 11, no. 2, pp. 161-168, 2012.
19. UjjwalMaulik, AnirbanMukhopadhyay, DebasisChakraborty, “Gene-Expression-Based Cancer Subtypes Prediction Through Feature Selection and Transductive SVM”, IEEE Transactions on Biomedical Engineering, vol. 60, no. 4, pp. 1111-1117, 2013.
20. Shuichi Kawano et al., “Identifying Gene Pathways Associatedwith Cancer Characteristics viaSparse Statistical Methods” IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 9, no. 4, pp. 966-972, 2012.
21. K.Ananthajothi, M.Subramaniam, Multilevel incremental influence measure based classification of medical data for improved classification, Springer, Cluster Comput. (2018). https://doi.org/10.1007/s10586-018-2498-z.

AUTHORS PROFILE

K.Ananthajothi (1983) is an Assistant Professor in the Department of Computer Science and Engineering at Misrimal Navajee Munoth Jain Engineering College, Chennai,(INDIA). He obtained his Master degree (M.E) in Computer Science and Engineering in Anna University in the year 2010 and Pursuing Ph.D in Anna University. His research focuses are Data Mining and Big-data. He published book title of Theory of computation.

M. Subramaniam (1974) is a Professor & Head for the Department of Information Technology at S.A. Engineering College affiliated to Anna University, Chennai, (INDIA). He obtained his Bachelor's degree (B.E) in Computer Science and Engineering from University of Madras (1998), Master degree (M.E) in Software Engineering and Ph.D from College of Engineering Guindy (CEG), Anna University Main Campus, Chennai -25 in the year 2003 and 2013 respectively. His research focuses are Computer & Mobile Networks, Cloud, Big-data and Software Engineering. He is an active life member of the Computer Society of India (CSI) and the Indian Society for Technical Education (ISTE). He has six Research scholars perusing Ph.D under his guidance. He published many research papers in reputed journals. He is also reviewer in IEEE- International Journal of Communication Systems.