Volume 8, No. 7, July – August 2017



International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

EFFECTIVE ANALYSIS OF BRAIN TUMOR USING HYBRID DATA MINING TECHNIQUES

B.V.Kiranmayee Associate professor Dept.of CSE VNRVJIET, Hyderabad,India Dr.T.V.Rajinikanth Professor Dept. Of CSE SNIST, Hyderabad ,India

Dr.S.Nagini Associateprofessor Dept.of CSE VNRVJIET,Hyderabad,India

Abstract: In the recent era there has been a rapid growth in the availability of medical databases and medical imagery in the past few years, and the uncertainty involved in the effective prediction of diseases from these databases made the research community to take up the challenges in this domain. The Central Nervous System of mankind being is mainly composed of Spinal Cord, Brain and Neurons. Largest part of brain is Cerebrum. The human brain throws good number of challenges to the research community. The Brain Tumour or the Intra Cranial Neoplasm is formed due to the irregular growth of cells in the brain. This sort of irregular growth of cells in brain damages frontal, temporal and parietal lobes thereby results in abnormal behaviour. Machine learning algorithms provide computers with the ability to learn without being explicitly programmed. The clinical brain data set was analysed effectively using machine learning algorithms and made conclusions on the results. In this paper we applied Hybrid Data Mining methods which in turn consists of Clustering, Classification andAssociation techniques and further analysed the results using some statistical techniques.

Keyword: Brain Tumor, Machine Learning Algorithms, Clinical brain data set, Hybrid Data mining Techniques, Statistical Methods

I.INTRODUCTION

There has been a rapid growth in the availability of medical databases and medical imagery in the past few years. The uncertainty involved in the effective prediction of diseases from these databases made the research community to take up the challenges in this domain. Brain tumor [1] is due to the uneven multiplication of cells in brain. Two types of Brain Tumors exists namely Primary(basic) and secondary, in case of Primary they originate in the brain itself whereas; in the case of secondary it spreads to brain from some other part of the body by metastasis and results malignant. Basic type of brain tumors will not affect or spread to the rest all body parts; they may be benign or malignant. These two types of tumors pose severe threat for life.Increased Intracranial Pressure occurs as there is limited space in skull. This ends up in reduced blood flow, edema and displacement. Further all healthy tissues are degenerated that act as controllers for all important functionalities in body. The next largest source of cancer is brain tumor which is enhancing the death rate in children and young adults. Brain tumor location specifies position of the tumor in the brain; it generates various symptoms, each of which can be related individual's functioning. Brain tumor not onlyaffects the Central Nervous System (CNS) and the Spinal cord, but many more vital parts of the body.

There are two possible ways of predicting or analyzing the Brain Tumors i.e. either by the application of Image Processing Techniques on MRI Images along with the classification or Data Mining Techniques application on Attribute information based data sets. The Data Mining Techniques were applied for Effective Analysis and also to arrive with desired results. Popular Data mining techniques are ARM (Association Rule Mining), k- Means clustering and decision trees. Categories of Data mining methods are Unsupervised, Reinforcement and supervised learning. Learning data in a supervised form is seen as learning based on a trainer which in turngenerates feedback to understand a new task. Feedbackoutcome is shown using a training set. Predictive modeling is used for analyzing existing database so as to govern the data's essentialcharacteristics. Usually for every model being developedtwo phases exists i.e., training and testing. These specializations are part of classification and enable in deriving valued prediction.

Algorithm C4.5 (J48): Quinlan in 1993 proposed this classification algorithm.By recursively partitioning the given data set this algorithm generates a decision tree. Depth-first method is utilized to build this decision tree internally. Algorithm approach is to split the data set that helps in selecting the best data set which enhances the information gain value.For every discrete attribute, one test with consequences of as numerous as the number of different values of the given attribute is measured. For every continuous attribute, binary tests are considered so that rest all possible distinct values of the attribute is addressed. To conclude, the value related to entropy gain is congregated by

applying efficiently all binary tests. This procedure is iterated for every continuous attribute.

Clustering:

This method enables decision making by grouping similar models. This gives a scope for understanding thestructure of data. The most concerned issue is in determining the total number of unknown clusters. One of the finest known examples related to unsupervised learning is clustering. The major goal of clustering lies in grouping similar data points by utilizing some metrics, the most commonly used metric being is Euclidean distance.

One of the most popular techniques for clustering is Kmeans. Reasons for the popularity of this technique include the absence of drawbacks of other types. For easy interpretation of results that could enhance the computational efficiency k-means technique usage is advantageous. The drawback of k-means include random choice of centroid locations at the beginning of the algorithm, handling of variables as numbers and the unknown count of clusters 'K'. By iterating and using a good initialization method the first drawback can be used instead of any good initialization method. Matching dissimilarity is a good solution for handling categorical data thereby eliminating the second drawback. The third drawback can be eliminated by using cluster validity index, this helps in deciding the count of clusters to be considered.

ARM (Association Rule Mining)

Association rule mining helps in discovering most related transactions. It finds the relationships between item sets built on their co-occurrence in the transactions.Particularly; it discovers the frequent patterns amongst the available item sets. For example, what are the most frequently purchased items in a shopping center.

Assume that there are a total of m items in the database; well defined as set of items I1 = $\{i1, i2, ..., im\}$. A transaction T1 is a set of all items such that T1 \subseteq I1.Consider the set of all transactions as D1. A transaction

T1 contains X1 if X1 \subseteq T1and as well as X1 \subseteq I1. Anassociation rule is a sort of implication taking the formX1 \rightarrow Y1 where X1 \subset I1, Y1 \subset I1 and at the same time X1 \cap Y1 = Ø. Two parameters are to be calculated for anyrule derived in this regard, one is confidence and theother is support. A rule R1 = X1 \rightarrow Y1 holds with the confidence value as c, if c% of the transactions o are part of D1 which also specifies that a transaction that has item X1 will also have item Y1 (i.e., c = P(Y1 | X1)). The rule R1 holds with support s if s% of the transactions in D1consists of both X1 and Y1 (i.e., s = P(X1 .Y1)).

II. RELATED WORK

V.P.GladisPushpaRathi et al. [1] stated that "feature extraction is a technique of catching image visual content in raw form and its storage in reduced form that facilitates decision making process, this type can be treated as pattern classification. This technique clubs features such as shape, texture, intensity and classifies the tumor as, gray matter, white matter, CSF, normal and abnormal area. One of the classification techniques SVM serves as measure for comparison between linear verses non-linear techniques. To reduce the number of features selected LDA and PCA techniques are used". RoopaliR.Laddha et al. [2] stated that "Exact detection of location and size of brain tumor plays a vital role in diagnosing it. An effective algorithm was proposed for brain tumor discovery which is based on morphological and segmentation operators. Quality of the image derived should be enhanced and it is subjected to morphological operators so as to detect the tumor. By using MRI and CT images redundant and complementaryinformation is captured by implementing wavelet based technique".

VarshaKshirsagar et al. [3] stated that "Simple Algorithms are used for detection of shape and range of tumor through brain images further computer aided methods like segmentation (detection) can be implemented to detect. Accuracy of tumor detection enhances on using segmentation techniques. Time for analyzing all such images also reduces. This entire process outcome would be detection of brain tumor with input being MRI along with its size and position. From the total area of the tumor derived from a specified cluster stage and the intensity of the tumor is defined".Punithavathy Mohan et al. [4] stated that "This deals with image segmentation, its enhancement, segment extraction and the classification of the Brain MRI. Proposed algorithm is CLAHE, k-means and ACO (AntColony Optimization). ACO is utilized for segmenting the image and k-means algorithm is utilized for classifying tissues as abnormal and normal from the brain MRI with very less time complexity and good level of accuracy". Jay Patel et al. [5] stated that "various clustering techniques like Fuzzy Cmeans, k-means Clustering have been used for different segmentation methods namely Thresholding, Region growing and Mean shift in MRI are reviewed".

Alan Jose et al. [6] stated that "Using fuzzy c-means, kmeans clusteringsegmentation carried out results in detection of brain tumor and its position. Fuzzy c-means performance is much better when compared with other techniques for segmentation. This process also enables in understanding the tumor stage, its severity and whether it is at a curablestate or not". Shubhangi S. Veer et al. [7] stated that"The global threshold and watershed segmentation areuseful techniques for the segmentation of MRI or CT scan images. The global threshold technique separates out diseased region by considering a single threshold point and further binary image is derived from the gray scale image. Watershed segmentation technique also separates out tumor region successfully from the brain tumor MRI region. From these segmented images it is possible to get the detail information about the tumor location, its shape and area of the extracted tumor regionis measured in terms of the number of white pixels in the segmented image".

Roshan G. Selkar et al. [8] statedthat "Brain tumor identification helps in the discovery of exact shape, size, location and boundary extraction of the tumor. The process has a total of three stages to identify and then to segment the brain tumor. Firstly the quality of image scanned has to be improved later the morphological operators were applied to identify the tumor in the imagescanned. After that edge detection operator is applied for boundary extraction and to find the size of the tumor". Gauri P. Anandgaonkar et al.[9] stated that "In order to locate tumor from the MR images derived different segmentation methods can be used. A method has beenproposed to carry out the same utilizing an algorithm called Fuzzy C-Means and then to calculate the area of tumor, this would enable in deciding on type of brain tumor". Rahil Malhotra et al.[10]stated that "segmentation of the brain tumor area for MR images was made and further it is subjected to image quantization process. This would help in applying the clustering process to understand various areas affected by brain tumor and finally with the ROI method detection of the brain tumor would reflect the segregated part of brain tumor".

A.K. Mohantly.et.al [11][14] have applied various techniques for the identification of cancerous tissues .Image mining is one of the technique through which we can mine the hidden data although the image is not in clear.By applying the image mining technique in testing point, space partition is used to categorize the image .S.W. Purnami.et.al [12] applied data mining technique" Multiple Knot Spline Smooth Support Vector Machine", for identifying the medical diagnosis issues majorly diabetes diseases and heart diseases.Shweta Kharya [13] has discussed about different data mining techniques which are used for the identification of breast cancer classification.N.H. Rajini.et.al [15][18] have projected majorly on two phases which are one of the feature extraction and the other is classification .In the first phase they have obtained the feature extraction which belongs to magnetic resonace imaging using discrete wavelet transformation and in the second phase they have used classification algorithms of feed forward back propagation artificial neural network and k nearest neighbor for automated diagnosis process

B.G. Prasad.et.al[16]have stated various data mining classifiers for the purpose of medical image classification where they have applied J48 decision tree and random forest for grouping the CT scan brain images.The projected developed system is mainly based on texture information of images.V.S. Tseng.et.al[17] projected image categorization methods by applying association rules on the image objects.The developed mehod is divided into phases .In the first phase construction of object hierarchy and in the second phase applying the multi level mining algorithms for identifying the image classification rules.A. H. Gondal.et.al [19] evaluated various techniques which highlights new outlook for brain tumor detection from magnetic resonance images.

III. PROPOSED SYSTEM

In the proposed system, at the beginning raw data setswere taken and refined so as to derive the refined data set by subjecting it to various methods of pre-processing.By which the noise and missing values were also addressed. Initially various classification techniques were applied and tested. The results derived are summarized and effectively analyzed for decision making. It was found that if initially clustering method k-means was applied, the resultant data set was taken as input for the decision tree classifier algorithm J48. The derived data set is subjected to Apriori Association algorithm. The derived results were analyzed further for effective decision making. Thus, this sort of Hybrid Data Mining technique helped in showing good results.

IV. EXPERIMENTAL RESULTS

From the Table-1 Depicts the results of 7 classifiers namely Nearest Neighbour with Generalization (NNge), Best-First Decision Tree (BF Tree), CDT: Decision tree the learner based on imprecise probabilities and the uncertainty measures, Hoeffding tree (VFDT), J48: it's a class for creating a pruned or un-pruned C4.5, LAD Tree and Random Forest. The results were depicted in Table-1 and Table-2. In Table -1 the parameters namely correctly MAE(Mean Absolute Error), classified instances, KappaStatistics, RMS(Root Mean Squared) error, RRSE (Root Relative Squared)Error, RA(Relative Absolute) Error, Time required to form the model and Time taken to Test the model values are depicted. It is observed from the Table-1 that two classifiers namely Random Forest, NNge did better in classification and next the LAD Tree did better than other set of classifiers. It is shown in Table-2 that the performance of 7 Classifiers in terms of parameter's namely FP Rate, TP Rate, value of Recall, value of Precision, value of ROC area, value of F-Measure and their accuracies.

From the Table-2 it isfound that the two classifiers namely Random Forest, NNge did better performance wise in classification and next the LAD Tree performed well when compared to other set of classifiers. To enhance the results of classifiers like J48 initially the data mining technique clustering namely k-means clustering method need's to be applied initially and then the resultant clustered dataset was subjected to J48 classifier so as to enhance the results. The results of clustered data set were shown in Table-3. The resulted clustered sets are namely,Meningesand other CNS (Cluster 0), Brain (Cluster1),Brain and the Nervous System (Cluster2).

Cluster 0: The age group centroid lies between 0-4. Most of the instances were clustered under this category. It is revealed from the Table-3 that Males Mortality rate is more when compared to Females mortality rate.

Cluster 1: The age group centroid lies between 40-44.Less number instances only i.e. 16% were clustered under this category. It is revealed that from the Table-3 that Males Mortality rate is more when compared to Females mortality rate.

Cluster 2: The age group centroid lies between 40 - 44.Less number instances only i.e. 16% were clustered under this category. It is revealed that from the Table-3 that Males Mortality rate is more when compared to Females mortality rate.It is observed that the Tumor types Brain and Brain andNervous System are behaving in the similar manner without much deviation. The Clustered data sets were shown in Fig.1.

It is observed that the J48 classifier results wereimproved from 67% before clustering to 100% after clustering. This is the advantage of Hybrid data mining and showed the performances of it in the Table -1 and Table-2. The resultant J48 decision Tree obtained on clustered data set was shown in Fig.2. The classification cancer type majorly depends on Mortality of person's Number attribute and is observed in Fig.2. The Clustered based classifier data set further subjected to AprioriAssociation Algorithm and 20 best Associations ruleswere generated and shown in Fig.3. It is observed from the association rules that Mortality of Persons Rate is depends on Mortality of Males Rate and mortality of Females Rate. The Fig.4 shows the 3-D visualization of Association rules in which the Antecedents and consequents were shown. B.V. Kiranmayee et al, International Journal of Advanced Research in Computer Science, 8 (7), July-August 2017,286-293

Classifier Technique	Nearest Neighbou rwith Generali zation(N Nge)	Best- First Decisio nTree (BF Tree)	CDT: Decision tree thelearner based on imprecise probabilities and uncertainty measures	Hoeffding tree (VFDT)	J48: Classf apruned or C4.5 Before Clustering	After	LADTree	RandomFor est
Correctly Classified Instances	96.3%	66.7 %	66.7 %	66.7 %	66.7 %	100	85.2%	96.3%
Incorrectly Classified Instances	3.7%	33.3 %	33.3 %	33.3 %	33.3 %	0	14.8 %	3.8%
Kappa Statistic	0.9444	0.5	0.5	0.5	0.5	1	0.7778	0.9444
Mean Absolute	0.0247	0.2222	0.2222	0.2222	0.2222	0	0.1286	0.1378
RMS error	0.1571	0.3333	0.3333	0.342	0.3333	0	0.2376	0.2147
Relative absolute	5.5556 %	50 %	50 %	49.5 %	50 %	0	28.9347%	31.0137%
Root Relative	33.3 %	70.7 %	70.7 %	72.5 %	70.7 %	0	50.4032 %	45.544 %
Timetaken to createthe	0.09	0.27	0.01	0.04	0.02	0	0.24	0.29
Timetaken to test the	0.01	0	0	0.02	0.02	0	0.02	0.03

TABLE-1: Various ClassifierResults and Analysiswhen applied on BrainTumorDataset

TABLE-2: Various Classifier PerformancebasedAnalysiswhen appliedon BrainTumorDataset

Dataset Name	Classification Technique	TPRate	FP Rate	Precision	Recall	F- Measure	ROC Area	Accuracy %
Brain	NNge	0.963	0.019	0.967	0.963	0.963	0.972	96.3
TumorData	BFTREE	0.667	0.167	0.500	0.667	0.556	0.833	66.7
set before	CDT	0.667	0.167	0.500	0.667	0.556	0.833	66.7
application ofK-means	Hoeffding tree	0.667	0.167	0.667	0.667	0.663	0.847	66.7
clustering	J48	0.667	0.167	0.500	0.667	0.556	0.833	66.7
erustering	LADTree	0.852	0.074	0.897	0.852	0.844	0.962	85.2
	RandomForest	0.963	0.019	0.967	0.963	0.963	0.998	96.3
Classificati on afterK- means clustering	J48	1	1	1	1	1	1	100

S.No.	Parameter/Attribut	ClusterNumber				
	e	Full Data(54.0)	0 (34.0)	1 (10.0)	2 (10.0)	
1	CancerType	Brain	Meninges andother CNS	Brain	Brain andthe NervousSystem	
2	Age	0–4	0–4	40–44	40–44	
3	I.M.NO	30.4444	11.8529	61.1	63	
4	I.M.R	8.5747	1.7857	19.8272	20.4051	
5	I.F.NO	22.1852	7.7059	45.6	48	
6	I.F.R	5.5775	1.2191	12.6622	13.311	
7	I.P.N	52.6296	19.5588	106.7	111	

B.V. Kiranmayee et al, International Journal of Advanced Research in Computer Science, 8 (7), July-August 2017,286-293

8	I.P.R	6.8867	1.502	15.7379	16.3431
9	M.M.N	23.7037	5.0588	55.1	55.7
10	M.M.R	6.9575	0.7338	17.4657	17.6098
11	M.F.No	16.1852	2.8824	38.6	39
12	M.F.R	4.4431	0.4343	11.1958	11.3205
13	M.P.N	39.8889	7.9412	93.7	94.7
14	M.P.R	5.5824	0.5869	14.0045	14.1451

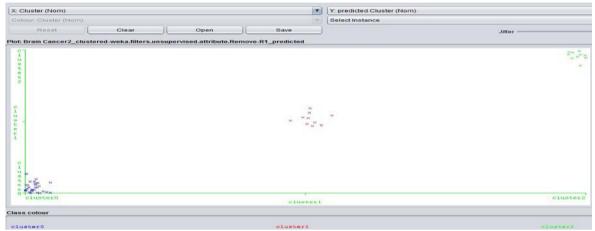


Fig.1: The Clustered dataset Visualization with Three cluster's

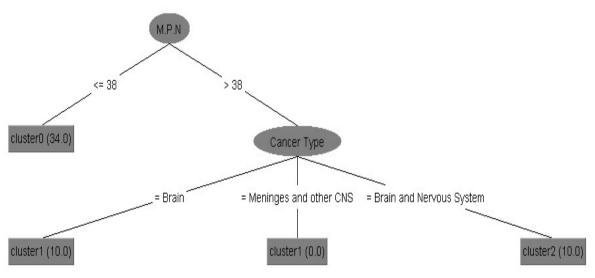
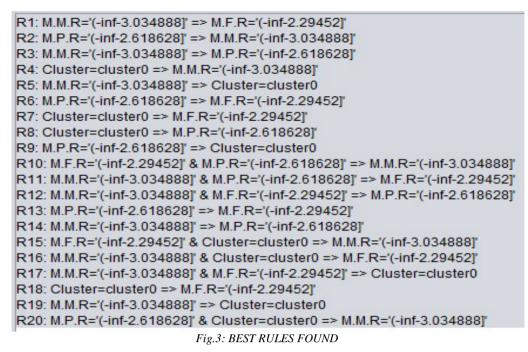


Fig.2: J48 DecisionTreeafterapplying it on theClustered dataset

B.V. Kiranmayee et al, International Journal of Advanced Research in Computer Science, 8 (7), July-August 2017,286-293



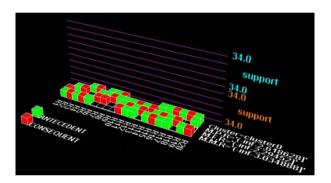


Fig.4: The3-DVisualization ofAssociation rules

V. CONCLUSION

By utilizing Hybrid Data Mining Techniques the results obtained for the Classifiers were improved drastically to a maximum extent. Details of which are shown through figures and the Tables.ThePerformances of Classifiers shown that NNge, Random Forest performed better in classification and next followed by LAD Tree when compared to other classifiers. The J48 decision Tree classifier results were poor when compared to the other classifiers.These results were improved on the Hybridization through clustering and Associations. The summary of the results shows that Males Mortality rate is more when compared to Females mortality rate and also that the Mortality of Persons Rate depends on Mortality of Males Rate and mortality of Females Rate.

REFERENCES

- V.P.Gladis, PushpaRathi, Dr.S.Palani,"Brain Tumor MRI Image Classification with Feature Selection and Extraction Using Linear Discriminant Analysis".
- [2] RoopaliR.Laddha, S.A.Ladhake, "A Review on BrainTumor Detection Using Segmentation And Threshold Operations", InternationalJournal of Computer Science and Information Technologies, Vol. 5 (1), 2014, pg.607-611.

- [3] VarshaKshirsagar, Prof.JagrutiPancha," Segmentationof Brain Tumour and Its Area Calculation", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 5, May 2014.
- [4] Punithavathy Mohan, Vallikannu. AL, B.R.ShyamalaDevi, B.C.Kavitha, "Intelligent Based Brain Tumor Detection Using ACO", International Journal of Innovative Research in Computer and Communication Engineering.Vol. 1, Issue 9, November 2013
- [5] Jay Patel and Kaushal Doshi, "A Study of SegmentationMethods for Detection of Tumor in Brain MRI", Advance in Electronic and Electric Engineering,Research India Publications, Volume4, Number 3 2014,pp. 279-284
- [6] Alan Jose, S.Ravi, M.Sambath, "Brain TumorSegmentation Using K-Means Clustering And Fuzzy C-Means Algorithms And Its Area Calculation", International Journal of Innovative Research in Computerand Communication Engineering Vol. 2, Issue 3, March 2014.
- [7] Shubhangi S. Veer (Handore), Dr. P.M. Patil,"Anefficient method for Segmentation and Detection of Brain Tumor in MRI images", International Research Journal of Engineering and Technology, Volume: 02 Issue: 09, Dec-2015.
- [8] Roshan G. Selkar, Prof. M. N. Thakare, "Brain Tumor Detection and Segmentation by using Thresholding and Watershed Algorithm", IJAICT, Volume 1, Issue 3, July 2014.
- [9] Gauri P. Anandgaonkar, Ganesh.S.Sable, "Detection andIdentification of Brain Tumor inBrain MR Images Using Fuzzy C-Means Segmentation", International Journal of

Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 10, October 2013

- [10] Rahil Malhotra, MadhuBahl, Harsimran Kaur, "BrainTumour Detection for Digital Images", International Journal of Advanced Research in Computer Science and Software Engineering ,Volume 3, Issue 9, September 2013.
- [11] A.K. Mohantly, S.K. Lenka, "Efficient Image Mining Technique for Classification of Mammograms to Detect Breast Cancer," International journal of computer science and communication technologies, vol. 2, no.3, pp.99-106, 2010.
- [12]S.W. Purnami, J.M. Zain, & A. Embong, "Data Mining Technique for Medical Diagnosis Using a New Smooth Support Vector Machine," Communications in Computer and Information Science, 2010, pp.15-27.
- [13] S. Kharya, "Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease," International Journal of Computer Science, Engineering and Information Technology, vol.2, no.2, pp.55-66, 2012.
- [14] P. Rajendran, M. Madheswaran, K. Naganandhini, "An Improved Pre-Processing Technique with Image Mining Approach for the Medical Image Classification," Second

BIBLIOGRAPHY OF AUTHORS

International conference on Computing, Communication and Networking Technologies, 2010, pp.183-187

- [15] N.H. Rajini, R. Bhavani, "Classification of MRI Brain Images using k- Nearest Neighbor and Artificial Neural Network", International Conference on Recent Trends in Information Technology, 2011, pp.863-868.
- [16] B.G. Prasad, A.N. Krishna, "Classification of medical image using data-mining techniques," Advances in Communication, Network, and Computing, 2012, pp.54-59
- [17] V.S. Tseng, M.H. Wang, J.H. Su, "A New Method for Image Classification by Using Multilevel Association Rules," Data Engineering Workshops, 2005, pp.1180-1187
- [18] Y. Zhang. and Z. Dong. "A hybrid method for MRI brain image classification", Expert System with applications, pp. 10049-10053, 2011
- [19] A. H. Gondal, and M. N. A. Khan, "A review of fully automated techniques for brain tumor detection from MR images," International Journal of Modern Education and Computer Science (IJMECS), vol. 5, no. 2, pp. 55, 2013

First author's Photo (3x4cm)	Mrs.B.V. Kiranmayee , currently working as Associate Professor in CSE at VNRVJIET,Hyderabad.she iscurrently pursuing Ph.D. degree inComputer Science and Engineering in JNTUH University. She has teaching experience of 18 years during which she has cultivated the teaching and administrative skills to excel in educational environment. Her research areas include Data Mining,Parallel computing and image processing
Second author's photo(3x4cm)	Dr. T. V. Rajini Kanth has obtained his Ph.D. degree in C.S.E branch from Osmania University, Hyderabad in July, 2008 and M. Tech. (C.S.E.) degree from Osmania University, Hyderabad in January, 2001.His specialization area is "Spatial Data Mining". His current research area interests includes Image processing, Data Warehousing & Mining, Spatial Data Mining, Web Mining, Text Mining, Machine learning and Robotic area, etc. His total teaching experience is 22 years. He is currently working as a Professor in CSE department, Sreenidhi Institute of Science and Technology (SNIST), Hyderabad
Thirth author's photo(3x4c	Dr.S.Nagini, working as an Associate Professor in CSE at VNRVJIET, Hyderabad.she has obtained her Ph.D. degree in ComputerScience and Engineering from Acharya Nagarjuna University in the year 2017. She has total teaching experience of 18 years. Her specializedarea of research is Data Mining