

Detection of brain tumor using K-nearest neighbor (KNN) based classification model and self-organizing map (SOM) algorithm

S. G. Raja, K. Nirmala

Abstract: Knowledge discovery is also known as Data mining in databases, in recent years that technique plays a major role in research area. Data mining in healthcare domain has noteworthy usage in real world. The mining method can enable the healthcare field for the enhancement of institutionalization of its administrations and become quicker with best in class technologies. Innovation utilization isn't restricted to basic leadership in undertakings, yet spread to different social statuses in all fields. In this paper a novel approach for the detection of brain tumor is proposed. The novel approach uses the classification technique of K-nearest neighbor (KNN) and for ignoring the error of the dataset image SOM (self-organizing map) algorithm has been used. Discrete wavelet transform (DWT) is used for transforming input image data set, in which RGB color of input data image has been converted into gray scale. Then it has been classified using KNN after that the error avoiding algorithm has been carried out. This will help to differentiate tumor cells and the normal cells. The presence of tumor in brain image is detected using parametric analysis by simulation.

Keyword: data mining, medical data mining, K-nearest neighbor (KNN), SOM (self-organizing map) algorithm, DWT (discrete wavelets transform)

I. INTRODUCTION

Recent decade has encountered a revolution in data accessibility and transfer of it through computer network. In the same spirit, many organizations of different industries have commenced to gather data relevant to their respective field. The database technologists are in search of efficient ways for storing, recovering and manipulating data. In order to learn and acquire knowledge from the data the machine learning developing method is essential. In strategic planning for research and development that the data is plays a vital role in knowledge Discovery database (KDD) and data mining (DM).Occasionally, the data can be plays an important key planning for innovative work which is stated as Data Mining (DM) and Knowledge Discovery in Databases (KDD). Data mining is an important area of research relevant to problem of discovering vague knowledge from real-time databases [1]. In enhancement of database technology, the input as well as output information of respective field has to be stored instantaneously.

Revised Manuscript Received on June 05, 2019

S. G. Raja, Research Scholar, Department of Computer Science , Vels University, Chennai, India

Dr. K. Nirmala, Research Supervisor, Department of Computer Science , Quaid-e-millath college for women, Chennai, India

One of the steps in KDD is data mining which contains analysis of information and computation of previously gathered information. Cluster analysis is an essential approach for database mining. It is either used as a unique tool to understand the distribution of an information index, e.g. to focus as well as support, data analysis and data preparation, or as a pre-handling adventure for various manipulations in the recognized groups.[2]

Data mining is a major approach to recognize the real, novel, potential significances in addition to sensible design in information. Data mining, the separation of the concealed perceptive data from tremendous database, is an extraordinary new advancement with potential to analyze crucial data in the data distribution center. The term data mining is referred to discovery of relevant and valued data from database. The mining methods can be applied on cognitive science to study human brain. Greater proportions of information are accumulated and stored in databases growing the necessity for profitability and convincing investigation techniques to make use of the data enclosed. The most necessary action for information investigation is group investigation. It helps the customer to know about the construction of data in database. Therefore, in the recent years the progress of enhanced group investigation has got a great deal of attention. [2]

In our human body brain plays a multipart organ. In human brain more than 100 billion nerves are located and overlapped. The identification of tumor in the brain is a difficult task because of its complex structure. An irrepressible development of cells in the brain is known as tumor. Malignant tumor and benign tumor are two kinds of brain tumor, these are smaller in size. Depends upon their growth pattern tumor are recognized. Benign tumor develops gradually within well-defined boundaries. This type of tumor can be completely detached by surgical procedure, such that it's not spread into any parts of brain, the spinal cord, or any other portion of the human body. The next type of tumor is known as malignant tumor, is rapidly-growing and affects the various cells of the healthy brain. In spinal cord this type of tumor is spread. It causes very harmful effects and may remain unprocessed. [3]

Brain tumors can be either benign (non-cancerous) or malignant (cancerous). Gliomas and malignant meningiomas, there are two

primary types of brain tumor or brain cancer. In malignant tumor, Gliomas is a familiar one. It includes different kinds of cells from which they existed: oligodendrogliomas, astrocytomas and ependymomas. Meningiomas started from the meninges, which surround the external part of the spinal cord and brain, which are tissues. The mainstream of meningiomas is benign and can be cured by surgery. It has numeral of brain tumors, with medulloblastomas, which develop from the primitive stem cells of the cerebellum and are most often seen in children. Primary and secondary tumors are the two kinds of malignant tumors. Auxiliary brain tumors as a rule begin in a few sections of the body and next metastasize to brain. The reasons for brain tumors incorporate head injuries, genetic syndromes, immune suppression, delayed introduction to ionizing radiation, electromagnetic fields, cell phones, or synthetic substances like formaldehyde and vinyl chloride. Manifestations of brain tumors incorporate determined headache, nausea and heaving, visual perception, hearing as well as discourse issues, strolling as well as parity troubles, identity changes, memory lapses, issues with cognizance and focus, and seizures. Meningioma is an assortment of tumor that creates from the meninges. The dura mater, arachnoid and Pia mater are the layers of meninges. Meningiomas are considered as benign tumors, with the 10% being atypical or malignant. Benign meningioma spreads slowly. A meningioma grows to a larger size before causing any symptoms. Other meningiomas develop very fast, and will have sudden growth spurts. There are no proper methods for manipulating the growth for a meningioma. Glioma is a tumor that arises from glial cells within the brain or spine. [4].

II. LITERATURE REVIEW

RuomingJin, Ge Yang, and Gagan Agrawal 2005, preformed an inside and out evaluation of arrangement of techniques, These techniques are investigated utilizing data mining algorithms, for example, fp-tree-based association mining, and apriori, k-means clustering, k-nearest neighbor classifier, and decision tree construction. The investigation completed in this paper accomplishes the accompanying: 1) Of the different techniques for parallelization of data mining algorithm, three techniques to be specific, full replication, optimized full locking, and cache sensitive locking can do very much contrasted with different strategies relying on machine and dataset parameters 2) The overhead of the interface is inside 10 percent in practically all cases 3) Good parallel effectiveness is accomplished for every one of the four data mining algorithms. 4) For phenomenal execution, the blend of different strategy swings to be basic if there should arise an occurrence of decision tree construction algorithm.[5] B.V.Kiranmayee, Dr. T.V.Rajinikanth, S.Nagini, 2016 presented an approach consisting of training and testing stage to detect the brain tumor. The functionality of proposed algorithm have been validated by constructing a prototype application. The results of developed prototype show that it can be combined to choice emotionally supportive networks in medicinal services area for enhancing

quality of services. [6] Ming-Ni Wu1 Chia-Chen Lin2 Chin-Chen Chang, In K- means clustering technique it utilized in color-based segmentation approach offers in order to recognized the tumor substance in magnetic resonance (MR).The vital thought of the proposed algorithm is to change the grey dimension MR image into a shading space image what's more it likewise achieves division for MR image by utilizing histogram-clustering and K-implies clustering. [7]

Saumya Chauhan[1],Aayushi More[2], RitumbhraUikey[3], Pooja Malviya[4], Asmita Moghe, 2017,suggests a median filter for pre-processing MRI brain images. Edge detection and the color based segmentation are being performed for segregation of lesion from image. The extraction schemes, such as gray level co-event lattice and histogram of situated gradients are utilized to extricate the different highlights of the pictures. All the removed highlights are put away in a debatable database. To this database, Instance based K-Nearest utilizing Log and Gaussian weight Kernels (IBKLG) classifier has been connected utilizing WEKA 3.9 instrument, to classify the tumor into benign or malignant. [8]. Samee Azad, Shaikh Anowarul Fattah, Naqib Sad Pathan 2016 proposes automatic method based on voxel statistics, for division of region of interest(ROI. The region of interest is a region encompassing the cerebrum tumor and its neighborhood. In the proposed technique, first conceivable applicant determination is performed using power qualities of tumor region in the T1 and FLAIR pictures of MRI documents. In the following stage, a cubic shaped 3D mean separating task is connected all in all volumetric data to get sifted volume are some irregular power conduct is relied upon to be dispensed with. At last, from the subsequent 3D FLAIR data, region of interest is removed dependent on combined dispersion capacity of power. It is discovered that the isolated region of interest offers real decrease of the general MRI volume without losing tumor data. The proposed ROI extraction plot is attempted on 20 real high survey tumor cases gained from a comprehensively used database and an exceptionally tasteful execution is gotten as far as segmentation precision, generally speaking volume decrease and computational time.[9] M. Murugesan, Dr. (Mrs.) R.Sukanesh2009 A programmed structure for expert recognition of brain tumors in electroencephalograms (EEG) signals utilizing artificial neural networks (ANNs). The ANN utilized in the projected structure is feed forward back spread neural system. Usually, the EEG signals will undoubtedly contain various historical facts about subject and devices impedances from the start with imperative data in regards to variations from the norm and brain action (reactions to specific stimuli). At first, adaptable separating is associated with confine the ancient rarities available in the EEG signal. In order to use spectral estimation the common features in the EEG signal



are segregated. Definitely, the fast fourier transform (FFT) analysis is carried out to extract the signal highlights covered in a varied group of noise. In this manner the perfect EEG information obtained is utilized as preparing contribution to the feed forward back engendering neural network. At the point when this prepared information is sustained with a test EEG signal, it adequately distinguishes the nearness of brain tumor in the EEG signal. [10]

SushmaV. Telrandhe, Divya Chikate, Pooja Banode,2015, explains an approach dependent on back propagations neural system procedure for the order of MRI picture. The approach is designed utilizing image enrichment, segmentation, registration, and characteristic identification and segregation techniques.

The morphological operations and thresholding process are considered during segmentation method. The training and experiment of MRI image is analyzed by back propagation neural network technique to recognize the presence of tumor. [11]

Neha Rani, Sharda Vashisth 2016 presents the work, in which to process the MRI images morphological and thresholding techniques are used. Feed-forward back propagation neural network is utilized to organize the presentation of tumor section of the image. By using the proposed method, high accuracy and less iteration for detection of tumor are achieved which in turn lessens the consumption time. [12]

Viktor Losing, Barbara Hammer and HeikoWersing2016 suggest the SELF ADJUSTING MEMORY (SAM) display for the K NEAREST NEIGHBOR (KNN) algorithm since KNN builds up a confirmed classifier inside the non-stationary information stream setting. SAM - KNN can manage unrelated idea point, for example, utilizing organically motivated memory models and their coordination. It very well may be effectively connected by and by since an enhancement of the Meta parameters isn't fundamental. The fundamental thought is to build committed models for the present and previous ideas and apply them as indicated by the requests of the given circumstance.

The different benchmarks, including fake streams with known drift includes just as genuine world datasets are generally assessed. Along these lines, we correctly include new benchmarks for exact assessment on different kinds of drift. The very focused results got from all examinations stress the robustness of SAM-knn just as its capacity to oversee heterogeneous drift idea. [13] WANG Huai-bin, YANG Hong-liang1,2XU Zhi-jian1,2 YUAN Zheng 2010, focus on improving intrusion detection rate and on SOM to give the accurate clustering results. Hence it Presents a new algorithm, in which SOM is used to gain generally clusters and focus of clusters, at that point, utilizing K-Means to refine the clustering in the SOM stage. The proposed algorithm is tested using KDD CUP-99 dataset. The test results illustrates that the proposed algorithm can achieve clustering accuracy and good stability of efficiency. [14]

III. RESEARCH METHODOLOGY

In this paper we suggest the novel move toward for the detection of brain tumor. Here we use the classification technique of K-nearest neighbor (KNN) and for ignoring the error of the dataset image SOM (self-organizing map) algorithm has been used. The image of input data set is transformed from RGB color into gray scale by DWT.

Then it is classified using KNN after that the error avoiding algorithm has been carried out. This will helps to differentiate tumor cells and the normal cells. Simulation results show the parametric analysis of the brain image shows if there is a presence of tumor. The architecture for proposed system is as follows:

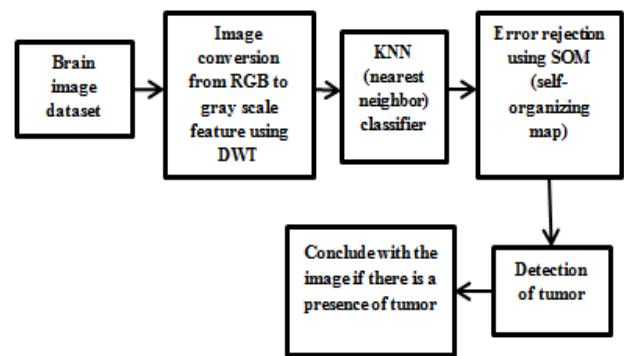


Fig.1 Block Diagram for proposed system.

A. Image conversion from RGB to gray scale feature using DWT

The gray level features are of two types i.e. edge mapping and energy. An edge mapping is a process in which fundamental implementation in image process, gives the contour of the object in an image. The result of edge mapping can trace the boundary of the object as well as the curve surface. Edge mapping is used for image segmentation. An energy level find by discrete wavelet change. The wavelet is amazing numerical instrument for highlight extraction of MRI image. The wavelet transform is used to give the information about the signal both in time domain and frequency. DWT (discrete wavelet transforms) is applied on image to get energy level.

H= undefined if max=min

$$\begin{aligned}
 H &= 60 \times \frac{G-B}{MAX-MIN} + 0 \quad \text{if } MAX = R \text{ and } G \geq B \\
 H &= 60 \times \frac{G-B}{MAX-MIN} + 360 \quad \text{if } MAX = R \text{ and } G < B \\
 H &= 60 \times \frac{B-R}{MAX-MIN} + 120 \quad \text{if } MAX = G \\
 H &= 60 \times \frac{G-B}{MAX-MIN} + 240 \quad \text{if } MAX = B \\
 S &= \left\{ \begin{array}{ll} 0 & \text{if } MAX = 0 \\ 1 - \frac{MIN}{MAX} & \text{otherwise} \end{array} \right\}
 \end{aligned}$$

V=MAX

The following steps are followed for conversing,



RGB color of image to gray scale feature using DWT

1. Initially RGB image is changed into YCbCr, follow by segregation of the Y, Cb and Cr mechanism.
2. Discrete Wavelet transformation is functional in the Y constituent, which results in 7 uneven substitute-images.
3. The components Cr and Cb are scaled at 25% of preliminary dimension, to extract Cb + as well as Cb- and similarly Cr+ as well as Cr- images. Optimistic/negative values of pixels are derivative from Cb and Cr components and others are equaled with 0.
4. The quarter from right base of left best sub-image is replaced with Cb-, the correct best sub-image with Cr+, the sub-image from left-base with Cb+ and sub-image from right-base with Cr.
5. DWT change is connected to the subsequent picture, to get the favored gray scale picture.

B. K-nearest neighbor Classification model

KNN is a simple classification method which works well for real time application. The training procedure is very easy and it sample includes class labels and set of tuples interrelated with that. This algorithm works for random number of module. Distance function is used by KNN classification model for mapping the samples with classes.

In order to calculate distance among the assumed test illustration X with that existing samples $y_1, y_2 \dots y_k$, the Classification process of KNN is used. The nearest neighbor around the test instance are identified and depends on the selection of neighbors, the majority neighborhood lecture is allotted to the test samples. The distance function is applied between the samples using Euclidean method or Manhattan method or Minkowski method. These methods are employed when the values are continuous. Depends on number of neighbor that the sample X probability is assigning. The probability of assigning a sample X to that of a class C is based on the number of neighbors considered, denoted as K.

$$\text{Probability of X to a class C} = \frac{\sum_{i=1}^k \text{distance}(c, c(y_i))}{k}$$

C. Self-organizing map for error rejection:

SOM algorithm undergoes the following stages to reject errors:

1. actualization– Introduce irregular qualities for the underlying weight vectors w_j .
2. Selection– Segregate an example preparing input vector x from the information space.
3. Identical– Compare the weight vector closest to that of the input vector and find the winning neuron $I(x)$. that is the least value of $d_{j_i, x_i, w_{ji}}, d_j(x) = \sum_{i=1}^D (x_i - w_{ji})^2$
4. Modernizing–Apply the weight refresh equation $\Delta w_{ji} = \eta(t) T_{j, I(x)}(t) (x_i - w_{ji})$ where $T_j, I(x)(t)$ is a Gaussian neighborhood and $h(t)$ is the education rate.
5. Continuance– Repeat Step2 until the feature map stops varying.

IV. PERFORMANCE ANALYSIS

In this section we discuss the results obtained using simulation. The performance analysis explains that the images of RGB color are transformed to gray scale using DWT and the parametric analysis of K-NN classification model and the error rejection using SOM.

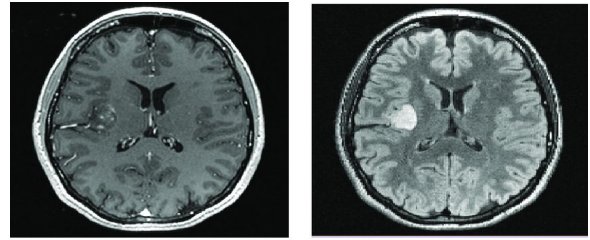


Fig.2 Brain Image dataset

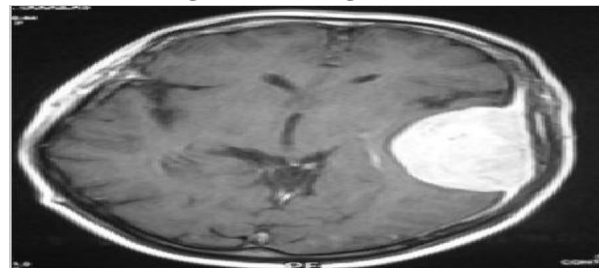


Fig.3 affected part of brain

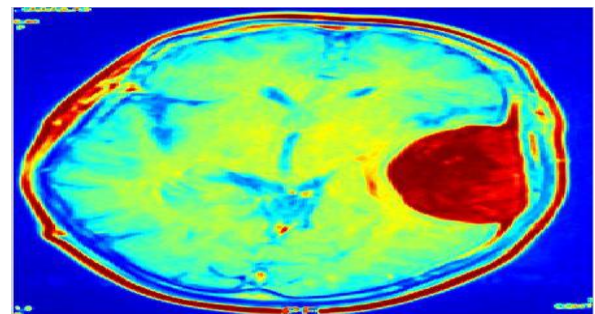


Fig.4 RGB colored Brain image

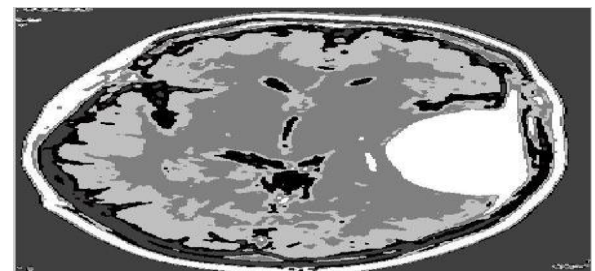


Fig.5 Gray scale image

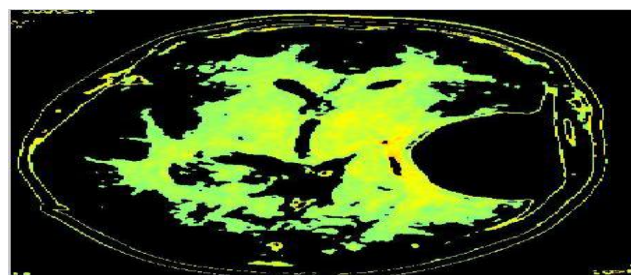


Fig.6 KNN classified image



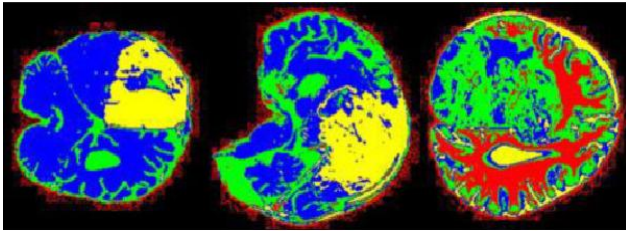


Fig.7 SOM error rejections in detected brain image

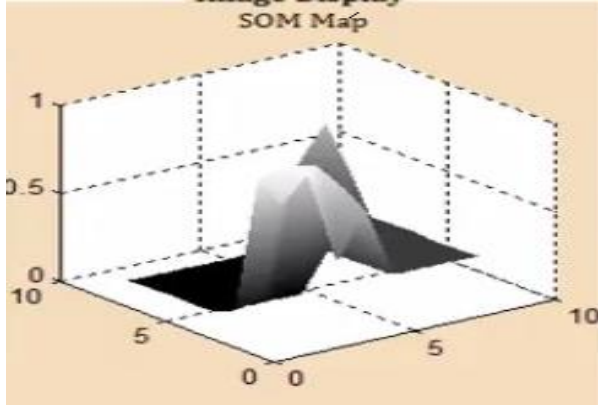


Fig. 8 SOM map for brain tumor detection

Fig.2 shows the brain image dataset, Fig.3 shows the affected part of brain without noise, Fig.4 shows the RGB colored brain image, Fig. 5 shows the gray scale image, Fig.6 shows the K-NN classified image, Fig.7 SOM error rejections in tumor detected brain image, and Fig.8 shows the SOM map for brain tumor detection.

Table-1 parametric analysis of K-NN and error rejection using SOM

Parameters	K-NN classification model	Rejection of error using SOM
Accuracy (%)	88.6	0.056
Sensitivity (%)	92.3	0.067
Specificity (%)	78.5	0.089

The table-1 represents the analysis of K-NN classifier and error rejecting algorithm using SOM of parameters such as accuracy, sensitivity and specificity.

V. CONCLUSION

In this paper suggests the novel method for the finding of brain tumor. In this manuscript, classification technique of K-nearest neighbor (KNN) and SOM (self-organizing map) algorithm for ignoring the error of the dataset image has been used. Using the proposed technique, the input image dataset, converts RGB color into gray scale using discrete wavelet transform (DWT). In next stage, it has been classified using KNN and later the error avoiding algorithm has been carried

out to differentiate tumor cells and the normal cells. From the experimental results it is clear that the MRI brain image has been classified using K-NN model and error rejected using SOM technique to detect the tumor in the brain image.

REFERENCES

1. Dr. Rajni Jain, "Introduction to Data Mining Techniques".
2. Shivangi Bhardwaj, May 2017, "Data Mining Clustering Techniques – A Review", IJCSMC, Vol. 6, Issue. 5, pg.183 – 186.
3. Shubhangi S. Veer (Handore)1, Pradeep M. Patil, | Dec-2015 , "BRAIN TUMOR CLASSIFICATION USING ARTIFICIAL NEURALNETWORK ON MRI IMAGES", IJRET: International Journal of Research in Engineering and Technology, Volume: 04 Issue: 12.
4. C.Ramalakshmi† and A.JayaChandran,May 2014, "Automatic Brain Tumor Detection in MR Images Using Neural Network Based Classification", IJCSNS International Journal of Computer Science and Network Security, VOL.14 No.5.
5. RuomingJin, Ge Yang, and Gagan Agrawal, Jan 2005, "Shared Memory Parallelization of Data Mining Algorithms: Techniques, Programming Interface, and Performance", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 17, NO. 1.
6. B.V.Kiranmayee, Dr. T.V.Rajinikanth, S.Nagini,2016, "A Novel Data Mining Approach for Brain Tumor Detection", 978-1-5090-5256-1/16/\$31.00_c IEEE.
7. Ming-Ni Wu1 Chia-Chen Lin2 Chin-Chen Chang, "Brain Tumor Detection Using Color-Based K-Means Clustering Segmentation"
8. 8.Saumya Chauhan[1], Aayushi More[2], RitumbhraiUikey[3], Pooja Malviya[4], AsmitaMoghe,2017 "Brain Tumor Detection and Classification in MRI Images using Image and Data Mining", ISBN 978-1-5090-4760-4/17/\$31.00©2017 IEEE.
9. 9.Sameez Azad, Shaikh Anowarul Fattah, Naqib Sad Pathan, 2016 "An Efficient Scheme for Detecting Region of Interest Encompassing the Brain Tumor from 3D MRI Data Based on Voxel Statistics", 2016 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE).
10. M. Murugesan, Dr. (Mrs.) R.Sukanesh, 2009, "Automated Detection of Brain Tumor in EEG Signals Using Artificial Neural Networks", 2009 International Conference on Advances in Computing, Control, and Telecommunication Technologies.
11. 11.Sushma V. Telrandhe, 2Divya Chikate,May-2015 "AUTOMATED BRAIN TUMOR DETECTION USING BACK PROPAGATION NEURAL NETWORK", International Journal of Soft Computing and Artificial Intelligence, ISSN: 2321-404X, Volume-3, Issue-1.
12. 12.Neha Rani, Sharda Vashisth, July 2016"Brain Tumor Detection and Classification with Feed Forward Back-Prop Neural Network",International Journal of Computer Applications (0975 – 8887) Volume 146 – No.12.
13. 13.Viktor Losing, Barbara Hammerand HeikoWersing, 2016,"KNN Classifier with Self Adjusting Memory for Heterogeneous Concept Drift", 2016 IEEE 16th International Conference on Data Mining.
14. WANG Huai-bin, YANG Hong-liang1,2 XU Zhi-jian1,2 YUAN Zheng, 2010, "A clustering algorithm use SOM and K-Means in Intrusion Detection", 2010 International Conference on E-Business and E-Government.