MRI Brain Tumour Detection & Image Segmentation by Hybrid Hierarchical K-means clustering with FCM based Machine Learning Model

[1]Sarita Simaiya, [2]Umesh Kumar Lilhore*, [3]Devendra Prasad, [4]Deepak Kumar Verma

*[1, 2, 3] Chitkara University Institute of Engineering and Technology,
Chitkara University, Punjab, India
[4]Chhatrapati Shahu Ji Maharaj University, Kalyanpur,
Kanpur Uttar Pradesh, India*
[1]umesh.lilhore@chitkara.edu.in, [2]sarita.simaiya@chitkara.edu.in,
[3]devendra.prasad@chitkara.edu.in,[4]deepak30057@rediffmail.com

*Corresponding Author: *Umesh Kumar Lilhore[2]*
***Email id:** umesh.lilhore@chitkara.edu.in*

ABSTRACT:**Introduction:** *A brain tumor tends to be a significant public health problem that affects many forms of living. To obtain an objective diagnostic to evaluate the best suitable medical plan, which derived data is used to classify the critical features within brain tumors i.e. position, size, and specific type.*
Background: *This research article developed a hybrid model called Hierarchical K-means clustering with Fuzzy c and Super Rule tree (HKMFSRT-Model) to introduce an asynchronous image recognition methodology for detection and identification of brain tumors throughout MRI images with their very beginning phases.*
Methods: *The suggested HKMFSRT-Model seems to be a hybrid of clustering k-means, Super-Rule-Tree, image acquisition utilizing patch-based system development, as well as to object counting approach. Several techniques need to have a predetermined fixed frame rate in advance to identify the patterns.*
Results & Discussion: *The proposed methodology uses a super-rule construct to formulate a plus-Rule-Tree to face the issue of misplaced patterns. Proposed method has accuracy result 88.9 %, and existing k-Means clustering method showing accuracy 85.4 %.*
Conclusions: *Beyond medical application, the procedure can also be incorporated into automatic treatment technologies and complex surgical applications.*
Keywords: *Brain Tumour, MRI image segmentation, Super Rule Tree, Machine learning, K-Means Clustering, FCM*

1. INTRODUCTION

The brain tumor seems to be an irregular area of slowly growing and multiplying tissue. This may originally come mostly in the brain, which is named a brain tumor tissue, but in other nearby neighborhoods at a future date transition to the brain, that is termed the metastatic brain tissue. Many as 140 forms of tumors have been categorized as benign and malignant. Tumor cells lead to serious health problems including extreme headache, weakness as well as stiffness. Each quality factor absorption frequency has been further analyzed to keep giving an organ cell perhaps a two-dimensional image. MRI processes have always been capable of generating photographs of various parts called slices inside the brain without overlapping certain anatomical structures as well as provide precise details regarding tumors along with precise location, structure as well as volume[1].

This information may help doctors but also practitioners show up for an objective tumor diagnosis as well as evaluate the correct treatment like surgery, and chemotherapy, as well as electro radiotherapy[2]. Due to the extreme immensely complex existence of so many of these data obtained, which therefore requires viewpoint throughout addition to trained and capable healthcare professionals, it is not an easy thing to manually retrieve essential health history from MR[3]. A hybrid model called Hierarchical K-means clustering with FCMA & Super Rule tree (HKMFSRT-Model) has been proposed throughout this research article. The proposed HKMFSRT-Model is a combination with clustering k-means, Super-Rule-Tree, image processing utilizing patch-based technique improvement, as well as to object counting. So the whole research is divided into phases representing introduction, related work, proposed HKMFSRT model, and experimental setup, and implementation, analysis of the results.

2. LITERATURE REVIEW

MR photos and videos are frequently used in the medical diagnosis including such detection of brain tumors, breast cancer, and so forth. That many investigators are working on brain tumor detection few of them seem to be as follows:

A hybrid approach with convex elevation data including watershed techniques, which involves estimating only the brain size via a three-dimensional transmission process, has been presented throughout the academic research paper[4]. Brain feature extraction coincides with data preprocessing which involves scraping non-cerebral tissues from the image using a method called skulls scraping, accompanied through strength convergence including noise pixel extraction. The skull stripping methodology was introduced in the research paper[5] that includes analyses to identify, density median filter, accompanied by that of the abolishment of limited connections links acquire brain to filter. The above strategies seem to be fairly effective in identifying the edge of just the brain which was massively complicated, computationally expensive as well as susceptible to both the possible consequence with deterioration as well as over fitting of brain tissues. In the research article[6] suggested a methodology focused primarily upon feature extraction to achieve a color wavelet correlation coefficients characteristic from fourth-order compositional measurements inside a percutaneous image. Some other method was proposed in[7] consists entirely of a centralized self-organizing map produced through removing noise present but instead classifying the basic tissue utilizing fcm-means clustering. The work suggested in[8] to tumor detection comprises three technologies: k-means clustering with such as watershed segmentation etc and optimized k-means clustering also with a genetic optimization technique.

The research article[9] integrates skulls strip with the fuzzy artificial neural network of Hopfield seems to be a strategy through tumor area diagnosis. Throughout specific, Berkley uses the technique presented throughout the paper[10] as well as supports vector transformer to enhance this same segmenting procedure besides features extracted from the segmented tissues. A method for all the classification of brain tumors, which covers image noise, clustering k-means, morphology operations, timing, and volumetric measurements was developed in a research study[11]. Through its experimental study [12] there seems to be a brain tumor identification strategy proposed via three phases: noise reduction, feature enhancement including clustering using k-means. A powerful strategy towards brain tumor detection includes morphological procedures that enhance an object intensity following wavelet transformation with segmentation & clustering k-means for tumor extraction. The brain features an extraction strategy that incorporates K-means clustering with genetic algorithms that were proposed by a research paper[13].

The research study tried to introduce an automatic segmentation technique[14] consists of three phases: pre-processing, clinically based neuronal network as well as post-processing. A research article[15] suggested a method for brain tumor classification using ant colony optimization (ACO). A further new method has been developed in the research paper[16] that increases the accuracy of both the brain tumor segmentation caused by differences in the tumor region heat transfer coefficients. This method was used to measure tumor shapes based on the thermal fluctuations by the Pennes bioheat formula including the canny edge detection process. Brain tumor segmentation concentrated on hybrid segmentation as well as texture features composed of evolutionary Viennese de-noising testing, morphological methodologies for difference can be attributed-cerebral cells along with k-means cluster centers also with multiple linear regression-driven kernel-based and efficient c-means imaging method[17].

3. MATERIALS

The technique HKMFSRT proposed in this research mainly comprises the following operational activities: clustering k-means, FCM, super rule tree, patch-based, tumor, and counting objects discussed in the following sections. **These include:**

3.1. Feature Extraction: The efficiency and accuracy of the proposed HKMFSRT model have been improved by eliminating different aspects even before hybrid K-means and FCM becomes produced. As we perform on just this strength, both existing clustering methods k-Means as well as the Fuzzy C-Mean clustering only can't generate amazing outcomes. In the very first phase, all these estimates rely upon around each pixel but we'd like to locate increasing pixel and just use the analytical details.

3.2 k-Means Clustering Method: The k-Means clustering seems to be an unsupervised clustering method which divisions each width of its cluster centroid dependent image, supported by the measurement of the distance around

each pixel intensity as well as the respective centroid. This method assigns increasing pixels according to the minimum distance rate toward a particular centroid. This method frequently adjusts its centroid by identifying the mean distance values of its centroid's allocated pixels. About the latest centroid, its distance values will be modified as well as the pixel being reallocated. Accurate k-means clustering has been achieved by reducing the variability inside of clusters, which seems to be the Total of squares inside each cluster (TSW) as well as optimizing the variability among clusters, which seems to be the Total of squares among clusters (TSB).

$$TSW(C, K) = \sum (X_i - C_k) pow2 \qquad \text{------- eq(1)}$$
$$TSB(C, K) = \sum N_j (C_i - X) pow2 \qquad \text{------- eq(2)}$$

Where: k= Numbers of successfully recognized clusters, X= Centroid total, N = total number of objects within cluster C.

A much lower the average TSW appreciate a much higher the inter-cluster unit associated with both the cluster configuration indicated so the greater the maximum TSB quality of such a cluster configuration, a much higher the segregation degree.

3.3 FCM: A Fuzzy c-means clustering also recognized as FCM has been one of the finest established techniques of something like clustering to somehow probably automatically reorganize the wide variety of products of large datasets and acquire precise result, though and it usually falls inside of local optimal solution.

Steps for the FCM algorithm:

1) *Select the cluster center at an arbitrary initial.*
2) *$D=[D_{ij}]$ (a Matrix), Here the initial value of Matrix D(0)will be set by using the equation (3)*
$$D_{ij} = 1 / \sum (|Z_i - MID_j| - |Z_i - MID_k|)^{2/m-1} \qquad \text{------------- eq(3)}$$
3) *Find out the MID(k) = [MID_j],Matrix, for k step all the centers vectors using the equation (4)*
$$MID_j = [\sum\nolimits^{N}_{i=1} Dm^m_{ij} * Z_i] / [\sum\nolimits^{N}_{i=1} Dm^m_{ij} \qquad \text{------------- eq(4)}$$
4) *Modified the U(k) , and the D(k+1) as by equation (5)*
$$D_{ij} = [1] / [\sum\nolimits^{Mid}_{k=1} (\,\|x_i - c_j\| / \|Z_i - Mid_k\|)^{2/m-1} \qquad \text{------------- eq(5)}$$
5) *Break the process when a min value for j will found otherwise set $\|D(k+1) - D(k)\|$ and*
$\| D^{k+1} - D^k \|$ else repeat step (3).

Additionally, we have quite a contribution matrix, although each pixel shifts more towards the cluster and seem to have a greater likelihood of detection as well as we can still have about the same clusters between items and that we can only say to be segmented properly

Where: $D_{ij} = Z_i$ is membership degree value towards the jth set of the cluster, $Z_i =$ i^{th} data for the set Z,
$MID_j =$ j^{th} cluster centroid, m= real number value of Fuzzy exponent (which is >=1)

3.4 Super Rules (SR) tree generation: A Super process is applied all through discipline; utilizing trans-cubic filtering, LR pictures will be first rise-sampled to almost the same size as that of the initial Input images. Even as an HR picture does have the same dimensions as that of the trans-cubic interpolated picture, this termed as H0 images. This same root node has to be initialized with all the other patch pairs obtained from the LR-HR training.

3.5 Patch-Based Image Processing: The initial image has been split further into tiny patches through patch-based image recognition, which has been depending on time as well as enables effective to deliver the correct input images. An MRI image has been evenly split into several patches of both the initial MRI image throughout the proposed method. Therefore detect more localize tiny tumors that cannot be identified throughout the image's regular scale; each patch will be enlarged upwards to multiple times its initial size.

3.6 Tumors Detection Using Object Counting: An Object count seems to be a strategy used mostly for finding as well as measuring objects in images. An object is identified by a cluster from a sharp pixel within a binary object that represents a related component. Brain tumors have been classified as particles, even though individuals occur when brightness in MRI images. In diagnose tumors, an initial MRI picture becomes translated into several binary items utilizing specific threshold variables accompanied through persistent attrition coupled through object counting till each object becomes equivalent to the number of objects within each binary image. Afterward, the images created being combined into another image.

3.7 Boundary Extraction: In this phase, we retrieved each tumor's boundary component either from the result of K-means; those latter cells become marked as either a non-tumor cell or a tumor cell. Then the forms of Brain tumors are curable so it could be likely here those genes will be partly destroyed because tumors will grow rapidly. The Fuzzy C-Mean method also provides tumor results in a positive experience case of partial inclusion. Although we transferred a K-means outcome with the Fuzzy C-Mean test if they're not tumor cells; whether yes then those cells are identified as tumors-like, all results are summarized. We have used a single-connected neighborhood area to retrieve tumor boundary lines.

3.8 Tumor Evaluation Process: Tumors of a brain simply aren't very adjacent to the skull. Individuals usually happen inside the human brain neuronal cells, and therefore bright particles appearing near the entrance to just the skulls are regarded to be false rumors. A distance measure between both the probable tumor as well as the highest surround edge of its skull has been used to classify and remove these objects. When it has been reported to be under a predetermined limit, the potential tumor becomes known to be a false tumor.

4. PROPOSED METHOD

The Proposed Hierarchical K-means clustering with FCM & Super Rule tree (HKMFSRT-Model) is based on the following steps:

Algorithm: HKMFSRT-Model for Brain tumor detection and segmentation

1) ***Clustering of the further Image sequence in major groups utilizing K-mean:*** *Probable tumor growing stage as well as darkened which include segments and sub-mean.*

2) ***Localization including strengthening of the brain surface:*** *With this procedure, like skull stripping wherein the external layer is eliminated, the underlying edge becomes reinforced and then used to prevent potential tumors.*

3) ***Euclidean distance threshold selection:*** *Towards the brain cells encircling the edge of both the skull as well as the brain cells to consider removing false tumors.*

4) ***Translating the initial MRI object through multiple pictures****: the very first picture has been computed by adding a sufficient know k-means cluster only with the largest twisted surround edge extending toward the center, then choosing the largest associated part and bridging the gap. A second image has been obtained by modifying an initial Image data using only a small kernel average filter and dividing that Filtered Picture by some initial Image data modified in strength.*

5) ***Dividing a converted picture into the patches:*** *Which are scaled close to multiple folds their original state average target intensifies tiny tumors to detect them over the next phase.*

6) ***Complete the full fuzzy clustering partitioning****: For data and measure the clusters generated during the process. Operate the fuzzy clustering architecture cluster analysis and then stop it completely when it generates the highest clusters.*

7) ***Perform initial clustering by K-means as well as FCM:*** *Over recorded data already analyzed in phase (4). Use centroid as even the original centroid throughout the K-means clustering method in just about any cluster created in phase (2).*

8) ***Detection among all potential tumors through binary for each major update:*** *Transformation utilizing multiple threshold values, including the application of object counting through erosion till the number of items becomes equal to just one accompanied by region production.*

9) *For the original MRI image repetition stages (5)–(8).*

10) *Repeat this step (5)–(9) beginning from an enables the data and choosing the subsequent tumors that have been identified in repeated steps at least once or twice.*

11) ***Removing false tumors****: Those are near to a skull's largest accompanying surface through including it for tumor evaluation.*

5. DISCUSSION

The proposed **HKMFSRT Model** has been simulated using a total of 1521 MR images and python programming language used for implementation and coding.

5.1 About MR image datasets: The initial MR image data set used for this work involves images collected from the internet Kaggle data collected which includes 1521 MR Tumour. The method has been tested utilizing 1521 real MRI images for both the detection of tumors of tiny size as well as large size. Throughout the addition of larger tumors, selected MRI images showed several early-stage tumors.

5.2 Implementations: Implementation started through the use of hybrid hierarchical k-means as well as the Fuzzy C-Mean method for the entire data set of MR objects to segment the entire images into a set key area: bright areas which contain potential tumors, moderate-intensity areas which include normal tissue as well as darker areas that don't include tissues and skulls sections. The whole process culminated throughout the detection of the largest surrounding point. To acquire a pixel intensity of brain cells as well as other parts including the skull, gaps in the biggest nearby edge have been covered. This same distance measure threshold with probable tumor form of the skull has been decided based on the maximum number of brain pixel resolution. A threshold seems to be directly proportional to the volume of the brain based on the brain tumor segmentation, as shown in figure 1.2 below.
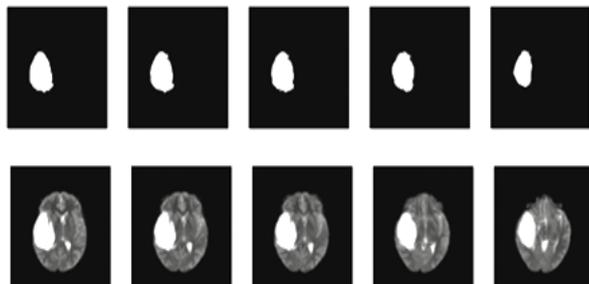

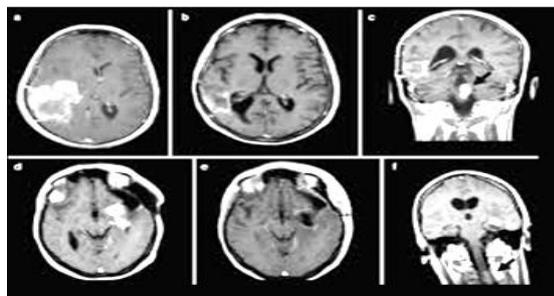**Figure 1.2 Tumor image segmentation**

Figure 1.3 MR Image Tumor detection

Another phase is usages through the morphological operations upon this original MRI image towards classifying tumors. Also at the bottom of the stage, the two separate images became combined through calculating the average, as described in figure 1.3.

5.3 Simulation Results & Comparisons
The proposed technique is sensitive enough to detect numerous large as well as small size and low-intensity tumors throughout MRI images, focused on the application as well as test results. A total of 1521 MR images has been used in this research, 60% for training including 40% for testing.

5.3.1 Comparison Parameters: The performance of the proposed HKMFSRT-Model has been further
Measured and evaluated by using various performance measuring criteria.
These performance measuring parameters mainly include the following-
➢ **Confusion Matrix:** An error matrix or confusion matrix is indeed a particular table structure which just enables mapping of the productivity of even a method, usually supervised learning .

N	0	1
0	TN	FP
1	FN	TP

Table 1.1 Confusion matrixes for Tumour detection

Where: Non tumour input=0, Tumour input=1, N= Total MR Images
TP= True positive, TN= True Negative FP= False positive, FN= False Negative
➢ **Accuracy:** This is the ratio of the participants correctly labeled to the whole pool of disciplines. Accuracy according to the following equation is the most intuitive.
$$\text{Accuracy} = [(TP+TN)/ (TP+FP+FN+TN)] \quad \text{-----------eq(6)}$$
➢ **Precision:** A precision is a fraction of what our system appropriately labels to all positively classified according to equation 7.
$$\text{Precision} = [TP/ (TP+FP)] \quad \text{-----------eq(7)}$$
➢ **Recall (or Sensitivity):** Recall seems to be the ratio of what our system correctly labels to all those who are correct in reality as per equation 8.
$$\text{Recall} = [TP/ (TP+FN)] \quad \text{-----------eq(8)}$$
➢ **F1-score:** The F1 Score needs to take both precisions and recall into consideration. It is just the avg or harmonic mean of accuracy as well as recall. An F1 Score is best if, as per equation 9, there seems to be some sort of compromise among precision & recall throughout the structure.
$$\text{F1 Score} = [2*(\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})] \text{-----------eq(9)}$$

➢ **Specificity:** A Specificity, as per Equation 10, is the correct negative labeled by the scheme to all those who are very healthy.

$$Specificity = [TN/ (TN+FP)] \qquad\qquad ----------eq(10)$$

5.3.2 Result comparison:

The proposed HKMFSRT-Model and Existing k-Means clustering model are implemented by using python programming in an anaconda environment. Following the simulation, the result has been calculated-

N=1521	0	1
0	TN=816	FP=138
1	FN=90	TP=477

Table 1.2 Confusion Matrix for existing k-Means Clustering Method

A Confusion matrix has been created for the existing clustering (k-Means clustering method), in which a total of 1521 MR images is used. Table 1.2 is showing the confusion matrix result for the existing k-Means clustering method. This matrix is clearly showing that the k-Means clustering method achieves 816 for TN, 138 for FP, 90 for FN, and 477 for TP.

N=1521	0	1
0	TN=843	FP=111
1	FN=63	TP=504

Table 1.3 Confusion Matrix for Proposed HKMGSRT Model

A Confusion matrix has been created for the Proposed HKMFSRT method, in which a total of 1521 MR images are used. Table 1.3 is showing the confusion matrix result for the Proposed HKMFSRT method. This matrix is clearly showing that the Proposed HKMFSRT method achieves 843 for TN, 111 for FP, 63 for FN, and 504 for TP.

Parameters	Existing k-Means clustering	Proposed HKMFSRT Model
Accura-cy	85.4	88.9
Precision	87.3	91.95
Recall	84.26	88.21
F1-Score	80.63	85.63
Specificity	85.33	88.22

Table 1.4 Experimental results for Existing k-Means Vs Proposed HKMFSRT
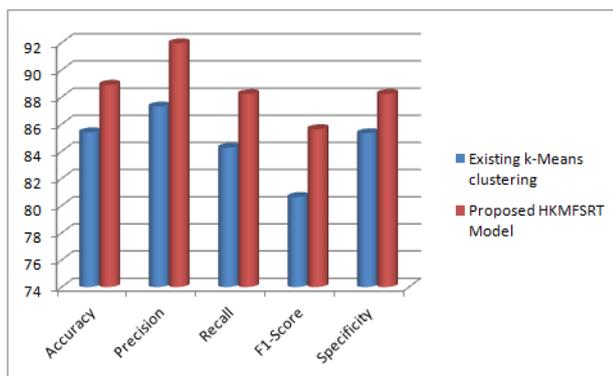
Figure 1.4 Experimental results for existing k-Means Vs Proposed HKMFSRT

The above table 1.4 and figure 1.4 is showing experimental results of 1521 MR images for existing k-Means clustering methods and proposed HKMFSRT methods. Existing k-Means clustering methods showing accuracy 85.4 % whereas the proposed method has accuracy result 88.9 % better than the existing k-Means clustering method. Similarly, k-Means clustering methods show precision %, recall %, F-1 score % and Specificity % simulation results as follows 87.3 %, 84.26 %, 80.63 %, and 85.33 %, and the Proposed method generates 91.95 %, 88.21%, 85.63 % and 88.22 % results for precision %, recall %, F-1 score %, and Specificity %.

6. CONCLUSIONS

The efficient segmentation of the MRI image seems to be an important phase in the care and treatment of brain tumors. The early detection of tumor cells enables doctors to assess the best treatment and thus improves patient chances of complete recovery. Throughout this article, we proposed a vision based technique to detect brain tumors in the initial stages from MRI images through means of the hybrid Model Hierarchy K-means clustered with the FCM & Super Rule tree (HKMFSRT-Model). The method has been validated and applied to utilize 1521 actual MRI images of the brain, which clinicians treated. Insurance which proposed method performs outstanding over the existing k-Means clustering system through experimental outcome analysis. In addition to large tumors, the

proposed HKMFSRT-Model has been capable of detecting early-stage tumors in Brain tumor MR data sets independent of its size, strength variation, and position.

ACKNOWLEDGEMENT

REFERENCES

[1] Mustaqeem, A., Javed, A., and Fatima, T., 2012. An efficient brain tumor detection algorithm using watershed & thresholding based segmentation. International Journal of Image, Graphics and Signal Processing, 4(10), p.34.

[2] Dong, H., Yang, G., Liu, F., Mo, Y. and Guo, Y., 2017, July. Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. In annual conference on medical image understanding and analysis (pp. 506-517). Springer, Cham.

[3] Patel, Mr Dharmesh, and Mr Umesh Lilhore., 2015. A survey on image mining techniques for brain tumor classification. International journal of engineering sciences & research technology, (pp. 592-595).

[4] Wu, M.N., Lin, C.C., and Chang, C.C., 2007, November. Brain tumor detection using color-based k-means clustering segmentation. In Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2007) (Vol. 2, pp. 245-250). IEEE.

[5] Ghanavati, S., Li, J., Liu, T., Babyn, P.S., Doda, W., and Lampropoulos, G., 2012, May. Automatic brain tumor detection in magnetic resonance images. In 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI) (pp. 574-577). IEEE.

[6] Iftekharuddin, K.M., Zheng, J., Islam, M.A. and Ogg, R.J., 2009. Fractal-based brain tumor detection in multimodal MRI. Applied Mathematics and Computation, 207(1), pp.23-41.

[7] Logeswari, T., and Karnan, M., 2009. An improved implementation of brain tumor detection using segmentation based on soft computing. Journal of Cancer Research and Experimental Oncology, 2(1), pp.006-014.

[8] Zook, J.M., and Iftekharuddin, K.M., 2005. Statistical analysis of fractal-based brain tumor detection algorithms. Magnetic Resonance Imaging, 23(5), pp.671-678.

[9] Dahab, D.A., Ghoniemy, S.S., and Selim, G.M., 2012. Automated brain tumor detection and identification using image processing and probabilistic neural network techniques. International journal of image processing and visual communication, 1(2), pp.1-8.

[10] Damodharan, S., and Raghavan, D., 2015. Combining tissue segmentation and neural network for brain tumor detection. International Arab Journal of Information Technology (IAJIT), 12(1).

[11] Bahadure, N.B., Ray, A.K., and Thethi, H.P., 2017. Image analysis for MRI based brain tumor detection and feature extraction using biologically inspired BWT and SVM. International journal of biomedical imaging, 2017.

[12] Ratan, R., Sharma, S., and Sharma, S.K., 2009. Brain tumor detection based on multi-parameter MRI image analysis. ICGST-GVIP Journal, 9(3), pp.9-17.

[13] Aher, Prashant, and Umesh Lilhore. Survey of Brain Tumor Image Quarrying Techniques , International Journal of Science & Engineering Development Research, ISSN:2455-2631, Vol.1, Issue 12, page no.111 - 115, December-2016,

[14] Logeswari, T., and Karnan, M., 2010. An improved implementation of brain tumor detection using segmentation based on the hierarchical self-organizing map. International Journal of Computer Theory and Engineering, 2(4), p.591.

[15] Vaibhavi, P., and Rupal, K., 2018. Brain Tumor Segmentation Using K-means–FCM Hybrid Technique. In Ambient Communications and Computer Systems (pp. 341-352). Springer, Singapore.

[16] Roy, S., Nag, S., Maitra, I.K., and Bandyopadhyay, S.K., 2013. A review on automated brain tumor detection and segmentation from MRI of the brain. arXiv preprint arXiv:1312.6150.

[17] Bhattacharyya, D., and Kim, T.H., 2011, April. Brain tumor detection using MRI image analysis. In International Conference on Ubiquitous Computing and Multimedia Applications (pp. 307-314). Springer, Berlin, Heidelberg.