



Special Issue of First International Conference on Science, Technology & Management (ICSTM-2020)  
**Detection of Brain Tumor Using Unsupervised Enhanced K-Means, PCA and Supervised SVM Machine Learning Algorithms**

Hari Prasada Raju Kunadharaju<sup>1</sup>, N. Sandhya<sup>2</sup>, Raghav Mehra<sup>3</sup>

<sup>1</sup>Research scholar of CSE, Bhagwant University, Ajmer & Vice-President, Wells Fargo Enterprise Global Services, Hyderabad, India.

<sup>2</sup>Professor, Department of Computer Science & Engineering, VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad, India.

<sup>3</sup>Associate Professor, Department of Computer Science & Engineering, Bhagwant University Ajmer, India. hari.kunadharaju@gmail.com<sup>1</sup>

**Abstract**

The brain tumor is an abnormal cell growth in the human body. To know which type of brain tumor it is and where is the exact location of it. We are using the MR image is a tomographic imaging technique. MRI is based on Nuclear Magnetic Resonance signals. A brain tumor is of two types 1. Benignant 2. malignant. Benignant belongs to I and II grade; this type of tumor is not active cells and have a low-grade tumor. It has a uniform structure. Malignant belongs to III and IV grades, this type of tumor are active cells and have a high grade. It has a non-uniformity structure. The initial phase Input MR image is transformed into a binary image by the Otsu threshold technique. The second step k-means segmentation process is used on binary images. Third step Discrete Wavelet Transform is used on segmented image for extracting the image and it reduces the large dimensionality by using PCA. It identifies the tumor by using Support Vector Machine classification it gives the final output of a brain tumor that normal or abnormal. The proposed paper experimented on the detection of brain tumors using classification algorithms dataset about BraTS dataset and compared with existing methodologies, and it is then proved that superior to existed.

**Keywords: Brain tumor, MRI, K-means, Segmentation, SVM.**

**1. Introduction**

Nowadays in the medical field, e-health care systems and information technology support clinical specialists to offer improved healthiness to the patient. Magnetic Resonance Imaging is the most powerful tool for producing the inside images of the human body, prominently in the brain. This technique is secure in comparison to other techniques like CT-scan and X-ray. MR brain images produce prosperous information for clinical diagnosis[1,2,3]. Briefly, the MRI is a Tomographic imaging method that gives the images of the core bodily and biochemical features of an entity from visibly measured nuclear

magnetic resonance (NMR) signs. MR imaging is created on the spectacle of NMR i.e., which was invented by Bloch and his colleagues at Stanford and Purcell and also his colleagues at Harvard in 1946. In 1952 Purcell and Bloch got the Nobel Prize in physics for this discovery.

A tumor is an abandoned expansion of a carcinogenic cell in any portion of the body, whereas a brain tumor is an abandoned expansion of a carcinogenic cell in the brain. According to the WHO and American Brain Tumour Association (ABTA), the brain tumor scale can be four categories. Based on these scales they can classify tumor types as either benignant or

malignant tumours. The benignant tumor comes under category I and II, which has homogeneousness in edifice and it doesn't cover lively cancer cells, and it is a low-grade tumor, whereas malignant tumours come under category III and IV, which have no uniformity in structure and it is a high-grade tumor. Patients with malignant tumours require continuous monitoring by MR Images or CT (Computed Tomography) scans every six to twelve months. The Brain tumor force impact any person at any stage, and its effect on the body may not be identical for every distinct. The Segmentation process is needed, for detecting infested tumor area from the medical imaging modes. In image analysis, segmentation is crucial. It is a technique for segmenting an image into various areas i.e., mutual and indistinguishable characteristics such as color, texture, boundaries, and shape. The segmentation of brain tumors comprises the procedure of segregating the tumor tissues from ordinary brain tissues such as enema, dead cells, and solid tumors, such as White Matter, Grey Matter, and Cerebrospinal fluid by using the MR images or any other imaging modalities [4, 5, and 6].

The paper aims to progress a computerized arrangement for enhancement, segmentation, and classification of brain tumors. The system can be used by neurosurgeons and healthcare specialists. The system incorporates image processing, pattern analysis and computer vision techniques and is expected to improve the sensitivity, specificity, and efficiency of brain tumor screening. The appropriate mixture and parameterization of the above stages empower the growth of assistant apparatuses that can aid in the premature diagnosis or the 24-hour care of the therapeutic measures.

## 2. Proposal Methodology

The main motto of this methodology is detecting the normality and abnormality of MRI brain scans. This sector discusses the block diagram of the projected architecture as shown in Figure 1

### 2.1 Image Thresholding

Thresholding is the humblest process of image segmentation. In this, the gray image is translated into a binary image. In the initial stage, the MR brain image is translated into a binary image by the Otsu Thresholding technique. The pixel values

larger than the threshold values are recorded as white and others are recorded as black.

### 2.2 Segmentation

Image Segmentation is a technique of segregating an image into a collection of connected sets of pixels i.e., into regions, linear structures, and 2D-shapes. Image Segmentation is the precise part of the clinical diagnosis. Depends on the image, segmentation can be a very sophisticated process. The segmentation of the infested regions of the MRI brains is attained by the subsequent stages: In the primary stage, the pre-processed MR brain image is transformed into a binary image by the Otsu Thresholding technique[7,8,9,10]. Then in the next stage, the segmentation process will be applied to the binary image. In this study, Enhanced K-means is used for effective segmentation of the brain.

The K-means machine learning technique is the best among the most popular partitioning method. In this method, the data is separated into a collectively k finite numeral of clusters and for the allocated each sample it gives the clustered index. In the case of massive data, the K-means clustering is highly deserved than the Hierarchical clustering. The clustering technique K-means is commonly applied for the unsupervised clustering procedure. Understanding the algorithm is easy and a stronghold when making it practice. But it does not provide valid accuracy. So, in this paper, we proposed Enhanced K-means i.e., selecting initial random centers with précising probabilities which increases the speed and also the correctness of the K-means.

Steps are shown below for Enhanced K-Means Algorithm:

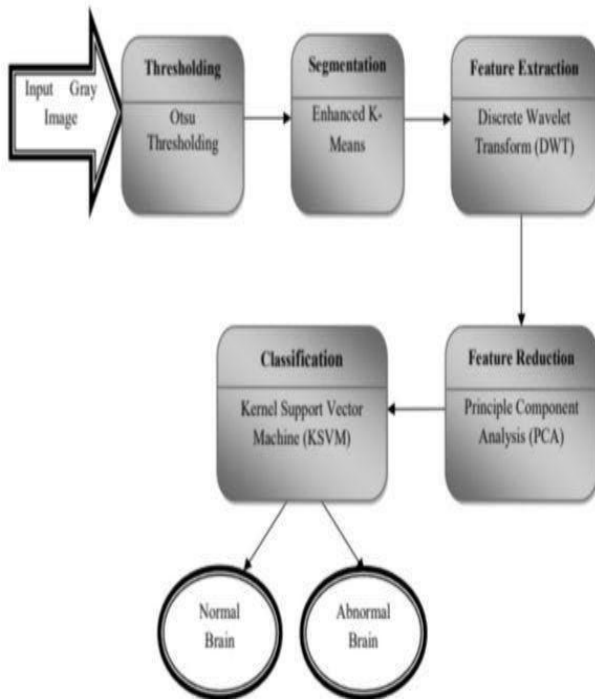
Phase 1: For the first center  $c_1$ , choose any data point randomly from the dataset  $k$ .

Phase 2: compute the distance from each data point to the center  $c_j$  in the dataset  $g=d(c_i, k_j)$ .

Phase 3: Determine the second centroid with

Phase 4: Determine the subsequent centroid randomly with the probability proportional to the distance from closest center to itself that you have already chosen from the dataset

Phase 5: Repeat phase 4 up to the prescribed number of clusters is framed after that follow the standard K-means algorithm

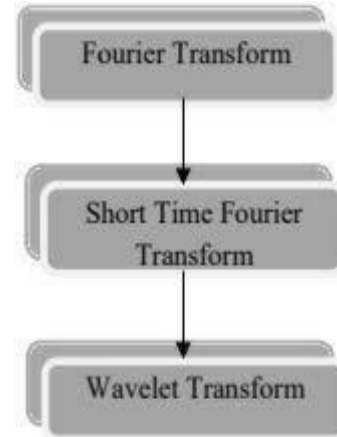


**Fig.1: Block diagram of the proposed Architecture**

**2.3 Feature Extraction**

It is the procedure of get-together perplexing level information from an image, for instance, shape, texture, color, and separation. Believe it or not, texture assessment is a huge limit of human visual understanding and machine learning structures. It is used effectively to improve the accuracy of the assurance system by picking obvious features. The primary task in feature extraction is pattern recognition i.e., takes the input pattern and gives it correctly to one of the desirable output classes[11,12]. This entire cycle is mentioned in two stages: one is Feature Selection, the other one is Classification. Highlight Selection is exceptionally fundamental in the whole cycle considering the way that the classifier won't be ready for seeing inadequately picked highlights. As per Lippman the rules to pick highlights are: "Highlights contains the data, which is critical to see classes, be difficult to irrelevant Variability in the information, furthermore be restricted in number, to allow, able assessment of discriminant limits and to keep the extent of arranging information required. The Fourier Transform (FT) is one of the orthodox tools for signal investigation i.e., which breaks the time domain signal into essential sinusoids of distinct frequencies, as follows, transforms the time domain signal to the frequency

domain. After all that, Fourier Transform has a very serious drawback i.e., it rejects the period info of the signal as shown in Figure 2.



**Fig.2. Development of the signal analysis**

The Gabor improved the Fourier Transform to analyze an insignificant sample of the signal at the time. This type of method is called the Windowing technique or Short-Time Fourier-Transform (STFT). It enhances a particular shape of the window to the signal. The STFT can be contemplated as an adjustment amongst the frequency info and time info. The STFT also affords info about both the frequency domain and the time domain. The WT (Wavelet Transform) presents a technique of windowing with variable size. The frequency information and time information of the signal will be preserved. The DWT (Discrete Wavelet Transform) is a robust orientation of wavelet transform by applying for diploid scales and positions. we are using 2D DWT. The DWT will enforce on 2D-images, which applies to each dimension separately. The image information will be interpreted in a simple hierarchical framework by using wavelets. In DWT, border distortion is the common technique, which is related to digital filters. When altering the image, at the edges the mask will enhance beyond the image, so the result is to wad the pixels outside of the images. In feature extraction, we have mainly 3 types. Those are Intensity, Shape, and Texture Features.

## 2.4 Feature Reduction

An enormous feature inflates the computation time and volume. The Curse of Dimensionality is a problem, it occurs due to the presence of a large number of dimensions or irrelevant features in the dataset. So, the dimensions or the number of features have to be reduced. The PCA is an effectual tool which reduces the dimensions of a data set, which consists of a huge number of interconnected variable quantity and possessing the contradictions. According to the importance, the data set will be transformed into a new set of ordered variables. It improves the computational efficiency and also maintains the classification accuracy.

## 2.5 Classification

It classifies the objects of an image into separate classes and gives the concluding step. Recently, many scientists have introduced many classification techniques. Predominantly it was categorized into two leagues i.e., one is Supervised and another one is Unsupervised ML. Concerning, the managed course of action system performs better than the independent portrayal methodology. Sponsorship Vector Machine (SVM) is a benchmark in the field of machine learning. It is a class of controlled learning. It gives high exactness, direct numerical agreement, and rich mathematical sensibility [11]. The SVM gives the maximum margin among hyperplanes that isolates one class from another class.

Kernel SVM (KSVM) recently, multiple kernel SVMs have been improved, grown rapidly. Those are very popular and effective. These kernel-based SVMs perform well when compared to conventional SVMs. The obtained solutions are unique, through avoiding the convergence to local minima possessed from the remaining probabilistic methods like neural networks. In this examination, we used the Gaussian part work for change. Using piece work the supportive request can be made, by the unit of non-straight models or data may get possible, where the non-direct models can be changed into a high dimensional future space. The Gaussian piece limit of nonlinear SVM moreover gives the ideal game plan of collection and hypothesis. we have different SVM classifiers applied and the results are compared. The training samples are chosen randomly from the dataset and

k-fold cross-validation is applied to justify the robustness of the proposed system. The k-fold cross-validation is very helpful to avoid over fitting.

## 3. Experimental Results

The dataset consists of 160 Brain MR Images with (140 abnormal and 20 normal). The dataset contains T2-weighted images of the brain; those are collected from the Harvard Medical School Website. We choose T2-weighted images instead of T1-weighted images because T2 Weighted images are higher in contrast and pleasant in perception in comparison with T1 Weighted images and positron emission topographies (pet) modalities like endoscopy, tactile imaging, and thermography.

**Table 1 (a): The Intensity Features for a few images**

Images	Mean	Standard Deviation	Variance	Skewness	Kurtosis
Image1	0.0031107	0.0897608	0.00804787	0.469022	7.32819
Image2	0.00250954	0.0897796	0.0080626	0.312064	6.23204
Image3	0.00348476	0.0897471	0.00798925	0.497891	6.52204
Image4	0.00341193	0.0897499	0.00805202	0.631759	7.68008
Image5	0.00458293	0.0896977	0.00806942	0.620389	6.5235

**Table 1 (b): The Intensity Features for a few images**

Images	Contrast	Homogeneity	Energy	Entropy	Correlation
Images1	0.208843	0.935159	0.7621	3.17346	0.199005
Images2	0.216073	0.93249	0.754802	3.31556	0.138167
Images3	0.251669	0.926743	0.740242	3.52392	0.073405
Images4	0.243326	0.934441	0.760611	2.9949	0.129391
Images5	0.243882	0.924625	0.731029	3.54839	0.107227

To validate the outcome of our algorithm, we used a dataset collected from the Harvard Medical School Website, which included sampling images of different patients. The sample preliminary results obtained from the proposed technique and that are depicted in Table 1(a) and (b) and Figure 3 (i) to (iv), Each step indicates how the extraction of the tumor is processed. The proposed PCA+DWT+SVM with the Gaussian Kernel method are better than the linear, poly, and

quadratic kernel SVMs.

The Gaussian kernel takes the exponential function form, which enlarges the distance from the samples to the proposed that the remaining can't reach. So, the GRB kernel can apply to other industrial fields also. We compared our method with the SVM different kernels GRB, LIN, and HPOL. PCA+DWT+KSVM (LIN) gives 95% accuracy, PCA+DWT+KSVM (HPOL) gives 96.1% accuracy and the proposed method PCA+DWT+KSVM (GRB) gives 97.8% accuracy. The GRB kernel performed best among the three kernels.

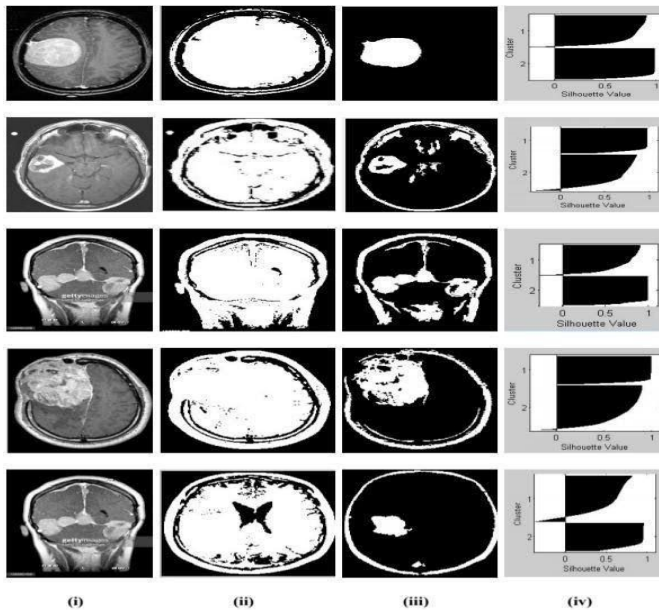


Figure 3 Segmentation results: (i) input MRI image, (ii) Otsu Threshold image, (iii) segmented tumor region, and (iv) silhouette graph for the clusters.

## Conclusions

The proposed paper is focused on accuracy and by using classification we get the result of the brain whether it is the normal or abnormal region. This paper involved feature extraction, feature selection, and classification for the MR image to get clarity of the image without any noise in the image. The proposed methodology on BraTS dataset images and also it is strong that the investigation for brain tumor detection is fast and accurate and also very much supportive to clinical experts and radiologists. In future work, to integrate all types of BraTS datasets and further progress the accurateness of the classification.

## References:

- [1].Nitta, Gnaneswara Rao, Nitta, Sneha." Dominant gray level based k-means algorithm for MRI Images," Journal of Health and Technology, Vol: 10, issue:1, pp: 281–287 (2020).
- [2].Andria, G., et al. "A new method to compare image quality in CT and MRI images." Medical Measurements and Applications Proceedings (MeMeA), 2011 IEEE International Workshop on. IEEE, 2011.
- [3].Scappaticci, Rosa, et al. "A feasibility study on microwave imaging for brain stroke monitoring." Progress in Electromagnetic Research B 40 (2012).
- [4].Kumari, Rosy. "SVM Classification an Approach on Detecting Abnormality in Brain MRI Images." International Journal of Engineering Research and Applications 3.4 (2013).
- [5].Gordillo, Nelly, Eduard Montseny, and Pilar Sobrevilla. "State of the art survey on MRI brain tumor segmentation." Magnetic resonance imaging 31.8 (2013).
- [6].Damodharan, Selvaraj, and Dhanasekaran Raghavan. "Combining tissue segmentation and neural network for brain tumor detection." Int. Arab J. Inf. Technol. 12.1 (2015).
- [7].Alfonse, Marco, and Abdel-Badeeh M. Salem. "An Automatic classification of brain tumors through MRI Using Support Vector Machine." Egyptian Computer Science Journal (ISSN: 1110–2586) 40.03 (2016).
- [8].Sudheera P., et al. "Detection of Dental Plaque using Enhanced K-Means and Silhouette Methods.
- [9].Surendar, A."FPGA based parallel computation techniques for bioinformatics applications", (2017) International Journal of Research in Pharmaceutical Sciences, 8 (2), pp. 124-128.
- [10]. V Ramakrishna Sajja., et al. "Segmentation of medical images using enhanced K-Means, Wavelet Transformations, and Morphological operations", CIE 46 Proceedings, 29-31 October 2016, Tianjin / China.

- [11].Parameshwari, Dasineni Sai, and P. Aparna. "An Efficient Algorithm for Textural Feature Extraction and Detection of tumors for a class of Brain MR imaging applications." Digital Signal Processing (DSP), 2014 19th International Conference on. IEEE, 2014.
- [12].Zhang, Yudong, Lenan Wu, and Shuihua Wang. "Magnetic resonance brain image classification by an improved artificial bee colony algorithm." Progress In Electromagnetics Research 116 (2011).processing 7.1 (2013)