A COMPARATIVE STUDY ON BRAIN TUMOR DETECTION USING IMAGE PROCESSING K-MEANS AND EM ALGORITHM

[1]Dr.S. Prabaharan, [2]Dr. R. Jegadeesan [3]B.Madhav, [4]K.Laxmi priyanka [5]T.Aishwarya

[1,2].Assoc.prof, [234]Under Graduate B.Tech Final Year Students-Computer Science and Engineering,

[1,2,3,4,5,6]Jyothishmathi Institute of Technology and Science, Karimnagar, India

Abstract :In this paper, we propose a comparative hybrid model for the detection of brain tumors based on the image processing, k-means,EM models . The K-Means algorithm is used to divide the image into K groups based on the distance between the pixels. Each group is grouped by the group center point,I.e., a centroid.In order to avoid inefficiency we use image processing. To remove the unwanted information (noise, film artifacts in MRI) in the original MRI image we use the pre-processing technique. We use the segmentation to extract a certain area of the image. We distribute an image to small parts and analyze information in a digital format.We use the extraction feature where some parameters are considered to extract features such as configuration, shape (shape), size and location of the image.Quality metrics such as PSNR is calculated to extract the features. With regard to the results obtained from extraction properties we perform the tumor classification process.We EM algorithm which is a general technique for maximum likelihood estimation or maximum a posteriori when there is a missing or unknown data.. Finally we compare the results obtained fron K-Means and EM algorithms and propose a best algorithm to detect the brain tumor .These techniques can then be applied to predict brain tumor detection at an early stage.

***IndexTerms* - Pre processing, brain tumor,feature extraction, classification, MRI,support vector machine.**

I. INTRODUCTION

A brain tumor is a group, or a block, of abnormal cells in your brain. Brain tumors can be cancerous (malignant) or non-cancerous (benign). Brain tumors are classified as primary or secondary. A major brain tumor develops in your brain. Secondary brain tumor, also known as transitional brain tumor, occurs when cancer cells spread in the brain of another organ, such as the lung or breast. Secondary brain tumors form the majority of cases of brain cancer. Diagnosing a brain tumor begins with physical examination and a look at your medical history. Physical examination includes a detailed neurological examination. These can include CT, and MRI scans of your head. Magnetic resonance imaging (MRI) differs from CT scan because it does not use radiation, and it does not It generally provides more detailed images of the brain structures themselves.

More than 700,000 Americans live with a brain tumor today. Approximately 80,000 people will be diagnosed with a major brain tumor this year. Approximately one-third (32 percent) of brain tumors and central nervous system (CNS) are malignancies. About 28,000 children in the United States are fighting brain tumors now. This year, nearly 16,000 people will die as a result of a brain tumor. Survival after diagnosis with primary brain tumor differs significantly by age, tumor type, location, and molecular markers. Brain tumors are the second most common type of cancer among children 0-14. More than 4,600 children and adolescents aged 0-19 will be diagnosed with a major brain tumor this year. Brain tumors and the central nervous system are the third most common type of cancer among adolescents and young people (ages 15 to 39) and the third most common cause of cancer death in this age group. The rate of incidence of central nervous system tumors in India ranges from 5 to 10 per 1,000,000 people with an increasing trend and represents 2% of malignancies. South India reported 15 years of experience involving 1043 patients.

Brain tumor detection of MRI images involves various stages such as processing, extracting parameters, chopping and grading. Classification is the last step in the brain tumor detection process where we implement the k-mean algorithm to divide brain images of MRI. The purpose of this study is to improve the accuracy of the detection of brain tumors using image processing tools and reduce the time taken to calculate the steps followed so that the image of the brain MRI can be identified as malignant or benign at the least time for the calculation .

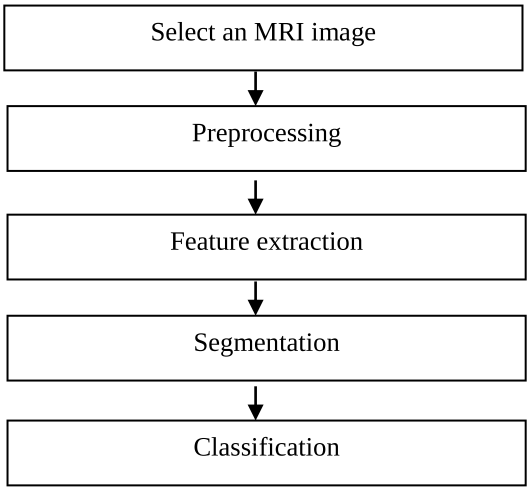

Fig1:steps for brain tumor detection

II. RELATED WORK

Brain tumors can be detected using image processing techniques by Gamage P. T [1] Pre-processing involves processes such as gradient conversion, noise removal and image reconstruction. Converting to a grayscale image is more common in preprocessing. After removing unnecessary artifacts, the image can be processed successfully. A image segmentation is a procedure for distributing image into small portions . Creates multiple sets of pixels within the same image. The partition makes it easier to further analyze and identify important information in a digital format. There are some criteria that are considered to extract features such as configuration, formatting (format), image size and image. With regard to results Retrieval of extraction features The tumor classification process is performed. K-clustering is used to divide data into specific number groups; K means Algorithm assigns each point of data into one of the K-clusters. The data point is set to the Terminal Cluster whenever the Euclidean distance is reduced. A K-mean aggregation algorithm is proposed by Ramish B. Kawadiwale1 and Milind propose a direct k-mean assembly algorithm by K. Alsabti, S. Ranka, and V.Singh [2] which is based on the regulation of vector mode so that all diseases that are closest to a given prototype efficiently pedicted.

In the first phase of the algorithm, we build a k-d tree to organize the pattern vectors. A root like this represents all patterns, while root children represent subsets of patients who have been fully contained in sub-areas (squares). The nodes are below These are the smaller funds. Mr. Cannago, David M. Mount, Nathan S. Netanyahu, Christine d. Biatko, Rotsselferman, and Angela Yu Wu [3] suggest that correlations based on k-means are closely related to a number of other aggregation and location problems. This includes the k-medians we option that describes the filtering algorithm. As mentioned earlier, the algorithm is based on storing multidimensional data points in the tree [kd] [4]. Each kd tree is associated with a closed box, called a cell. The root cell is the vector box for the set of points. If a cell contains only one point (or less generally less than some small stability), it is considered a sheet. Otherwise, it is divided into two hyper rectangles by a hyperplane orthogonal axis. The result is the original cell children, resulting in a binary tree.

In this study, digital image processing techniques are important for detecting brain tumor through MRI images. The Previous processing techniques include different methods such as filtering, contrast enhancement, and edge detection to smooth the image. Images that have already been processed are used for post processing processes such as; threshold, graph, segmentation and morohological, which is used to enhance images Which is proposed by Arti Gujar , Prof. C. M. Meshram .

We offer preliminary design and experimental results for the way cancer images are tracked to MRI images (some stock images) that use a color-shifted fractionation algorithm with K-mean assembly technique. This method can solve the problem of finely defined pest organisms in MRI image by adding a color-based fragmentation process. The main idea of a color-coded defragmentation algorithm with K-means is to resolve the MRI image by converting the grayscale image image to a colored space image and running the image that is marked with the cluster index which is proposed by li-hong jaung "MRI brain lesion image detection based on Color Converted K-Means Clustering Segmentation“.

This provides a general approach to calculating estimated frequency of probability when observations can be presented as incomplete data. Since each frequency of the algorithm consists of an expectation step followed by a maximization step, we call it an EM algorithm. When the full basic data comes from an exponential family whose maximum probability estimates are easily calculated, each maximizing step of the EM algorithm is easily computed. The term "incomplete data" in its general form refers to two sampling spaces fromY to X and X and one from X to Y. We refer to x as complete data which is proposed by A.P. Dempster; N. M. Laird; D. B. Rubin

Brain tumor fragmentation is an important task in the treatment of medical images. Early diagnosis of brain tumors plays an important role in improving treatment possibilities and increases the survival rate of patients. The manual division of brain tumors to diagnose cancer, through a large amount of MRI images created in the clinical routine, is a difficult and time-consuming task. There is a need to divide the image of an automatic brain tumor. First, an introduction to brain tumors and ways to segment brain tumors. Then, modern algorithms are discussed with a focus on the modern trend of deep learning methods. Finally, an assessment of the current situation is presented and future developments are addressed to standardize the methods of brain-based MRI brain segmentation in the daily clinical routine which is proposed by Ali IúÕna, Cem Direko lub , Melike ùahc " Review of MRI-based brain tumor image segmentation using deep learning methods ".

There is now a wide variety of image-splitting techniques, some of which are general-purpose and some are designed for certain categories of images. These techniques can be categorized as follows: spatial clustering associated with the measurement area, one-link area growth schemes, hybrid link area growth schemes, RTP growth schemes, spatial aggregation schemes, and division and integration schemes. In this research, we define each each of the major categories of image segmentation techniques and describe several specific examples for each category of algorithm. We illustrate some techniques with examples of the divisions that are implemented on the ground which is proposed by Robert M.Haralick and Linda G.Shapiro

Collation is a well-known method for extracting data that is used to group data elements together on the basis of a similarity property. Partial cluster algorithms get a single partition of data instead of a cluster structure. K-assembly means a shared approach. But one of its drawbacks is to randomly select midpoint points because of the algorithm that has to repeat the number of times. The hierarchical group creates a collective hierarchy, in other words, a group tree, also known as a dendrogram. Hierarchical groups can be further grouped into a group (bottom-up) and a top-down section. The widely used hierarchical aggregation algorithm is CURE (Assembly Using Represents Presentations) and BIRCH (heterogeneous multiplication and clustering using hierarchies) which is proposed by Kedar B. Sawant

III. RESEARCH METHODOLOGY

Brain tumors can be detected using image processing techniques. Preprocessing involves operations such as grayscale image conversion, noise removal, and image re-creation. Switching to the grayscale image is the most common exercise practice. After removing unnecessary artifacts, the image can be processed successfully. Image fragmentation: "Image splitting" is an image distribution procedure to small parts. Creates multiple sets of pixels within the same image. The partition makes it easier to further analyze and identify important information that forms a digital image.There are some criteria being considered for extracting features such as configuration, format (format), image size and image. With respect to the results obtained from extraction features the tumor classification process is performed.One of the most simple, fast, effective, unattended and popular assembly algorithms that solve the known clustering problem by separating data into K groups is the K-Means Clustering algorithm. The K-Means algorithm is used to divide the image into K groups based on the distance between the pixels. The K-Means algorithm is an iterative, statistical, non-deterministic algorithm, and a digital aggregate. Often used for simplicity of implementation and speed of convergence.It provides high quality packages by looking at the low level of the desired account. This method mainly focuses on reducing the total square spaces between all the pixels and the center of the cluster.

The K-Means algorithm relies on the similarity indicator or the difference between pixel groups. Each group is grouped by the group center point, ie, a central point. It's just a way to set points (a set of points). In this algorithm, the numbers of K points are predefined in general and K determines a number of groups. This is a variable algorithm that can be used for any type of assembly. It combines pixels based on distance, using the Euclidean space square(Alsabti et al., 1997; Lalaoui et al., 2015).The main objective of this algorithm is to divide the N points in the D-dimensional space into K blocks so that the distance from the pixel points of the image to the cluster decreases. Among assembly combinations that are based on the reduction of the function of the official goal, the most widely used and widely studied is clustering. Given the range of n data points in real d-dimensional space, Rd, and k integer, the problem is to determine a set of k points in Rd, called centers, so as to reduce the average square distance from each data point to the nearest center.The K-Means assembly algorithm has been modified and proposed in different ways according to needs. In the k-means algorithm, to avoid unnecessary comparisons between data points and means by comparing means with each other as well as to avoid overhead by sorting methods, Phillips (2002) introduced two modifications. Improves run time without changing the output, called the resulting algorithms meaning "means and means of sorting". Li-Hong Juang and Ming-Ni Wu (2010) presented the K-method of collection technique using the color-deflection algorithm converted to MRI images.The basic idea of this method is to convert the MRI image into a segmented color image. This method is able to follow the tumor and help doctors to distinguish the micro-lesion and the area.

K-Means Clustering Algorithm Initially, the central point are randomly assigned and pixel points are grouped at central points. Each pixel of the image is grouped to form a group, and the average of each group is calculated. If the mean value of a block is equal to the value of this central composite, no changes will be made or the value of the central point changes to the

average value of the block.After that, each pixel point of the image calculates the distance from the pixel point of the image to its own point point. If the pixel of the image is close to its center point, no changes are made, but if the pixel point of the image is not close to its center point, move it to the nearest central focal point. Again, the average value for each group must be calculated. Repeat the above process until the values of the center point and pixels of the cluster image become fixed
K-means can be used in a large set of data (images) and if K is small, it will work quickly. If the clusters are spherical, they may produce compressed clusters of pyramidal clusters. The main problems are the number of clusters K, and the different initial conditions lead to different results. The pixels away from the center pull the centers away from the optimal position.

Expectation maximization algorithm

The EM algorithm is a general technique for maximum likelihood estimation or maximum a posteriori when there is a missing or unknown data.EM algorithm is used for estimating the hidden ground truth and expert segmentation quality parameters from a collection of segmentations by experts. It is necessary to take the derivatives of the likelihood function with respect to all the unknown values and the parameters for finding the maximum likelihood solution. EM algorithm is a region and probabilistic based method that deals with the hidden data problems. It has the extremely fashionable property that each of its iterations will raise the posterior-likelihood. This algorithm can be used in two situations. The first occurs, if there is a problem with the data indeed, because of having missing values or limitations of the observation process and the second occurs during the optimization process. In EM segmentation algorithm, only one cluster center value is assigned initially and other cluster center points are assigned with respect to certain probability condition.

EM Segmentation Algorithm : This algorithm considers two parameters, namely observed data space and unobserved data space. Observed data are directly measured from the images and unobserved data (latent data or missed data) are the influence of observed data, but unobserved data are not small enough to measure.The E-step computes the expectation of the likelihood by including the missed variable as if they were observed. M-step, computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found on the previous E-step Until the clusters value converges, the E-step and the M-step are performed iteratively.The maximizing likelihood function is expressed in the terms of x, z, and θ.Here x is the observed data, z is the unobserved data and 'θ' corresponds to the set of all parameters of the distribution.

IV. CONCLUSION AND FUTURE WORK

The purpose of the study was to propose a new method for the diagnosis of brain tumor in its early stages using images of MRI. Among the solutions available, the solution was found that applied an intermediate candidate for pre-processing, k-means for segmentation, first statistical features, text extraction features, and Vector Machine for the most efficient classification.

REFERENCES

[1] Gamage P. T. - 144056U Under the supervision of: Dr. Lochandaka Ranathunga, "Identification of brain tumor using image processing techniques." ResearchGate 2017

[2] K. Sveti, S. Ranka, and Singh. K-Means efficient aggregate algorithm. http://www.cise.ufl.edu/ ranka /, 1997

[3] Tapas Kanungo, Senior Member, IEEE, David M. Mount, Member, IEEE, Nathan S. Netanyahu, Member, IEEE, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu. -Means Aggregation Algorithm: Analysis and Implementation "IQ, VOL. 24,. 7 July 2002.

[4] Ayakika Martinez, IO, Martinez-Trinidad, JF and Ariel Carrasco Ochoa, J 2006, "K-Means Conceptual Algorithm Based on Complex Features", Proceedings of the 11th American Iberamic Conference on Progress in Pattern Recognition and Image Analysis and Applications, p. . 491-501.

[5] Arti Gujar, Prof. C. M. Meshram. "Brain Tumor Segmentation and Classification":A Review. 2018.

[6] Li-Hong Juang & Ming-Ni Wu 2010, "MRI brain lesion image detection based on color-converted K-means clustering segmentation, Measurement ', vol. 43, No. 7, pp. 941-949.

[7] "Maximum Likelihood from Incomplete Data via the EM Algorithm" A. P.Dempster; N. M. Laird; D. B. Rubin Journal of the Royal Statistical Society. Series B (Methodological), Vol. 39, No. 1. (1977), pp. 1-38.

[8] Ali IúÕna, Cem Direko lub , Melike ùahc " Review of MRI-based brain tumor image segmentation using deep learning methods " 12th International Conference on Application of Fuzzy Systems and Soft Computing, ICAFS 2016, 29-30 August 2016.

[9] Robert M.Haralick and Linda G.Shapiro "Image Segmentation Techniques"Machine Vision International 325 E.Eisenhower, Ann Arbor, MI 48104.1985.

[10] Kedar B. Sawant "Efficient Determination of Clusters in K-Mean Algorithm Using Neighborhood Distance" International Journal of Emerging Engineering Research and TechnologyVolume3, 1 January 2015.