

# K-Means Segmentation and Normalized Histogram: An Effective Method for Detecting Brain Tumor from MRIs

K. Pugazharasi<sup>1\*</sup>, K. Sakthivel<sup>2</sup>, R. Vijayarajeswari<sup>3</sup>, N. Pushpa<sup>4</sup>

Submitted: 16/03/2024 Revised: 30/04/2024 Accepted: 05/05/2024

**Abstract:** In the medical field, brain tumors are evaluated using a technique known as magnetic resonance imaging or MRI. To eliminate the additive noise which is found in MRI images, including Gaussian, Salt & Pepper, and Speckle noise, this study employs a technique to examine and classifies image d-noising filters including Mean, Adaptive, Minimizing, UN-sharp masking filter and the Gaussian filters. The DE-noising efficiency of each method is investigated using PSNR and MSE. The effective brain tumor segmentation utilizing the normalized histogram and the K-means clustering algorithm is shown as a novel method. Support Vector Machine (SVM) is to provide accurate forecasting and classification.

**Keywords:** Segmentation, Histogram, MRIs, Clustering

## 1. Introduction

### 1.1 Background

Diagnostic imaging is becoming a vital tool in modern medicine. Clinical practice makes extensive use of Computed Tomography (CT), nuclear medicine (MRA), MR imaging (MRI), and X-ray angiography (XRA). These pictures add to the information we know about the patient. The essential element of automated radiological diagnosis systems is vessel segmentation algorithms. Different segmentation techniques are used depending on the available imaging modalities, application domain, automatic or semi-automatic nature of the procedure, and other unique aspects

There isn't a single segmentation technique that can remove the vasculature from every type of medical image. While some strategies use linked component analysis after thresholding, a pure ferocity-based pattern recognition methodology, other techniques use explicit vessel modelling to extract the vessel outlines. Some

segmentation techniques may require the processing of images before the segmentation algorithm, depending on the picture quality and general image artefacts like noise. Conversely, some methods use post-processing to fix the issues brought on by excessive segmentation.

Some segmentation techniques may require image processing before the segmentation algorithm, according to the image quality and other generic image artefacts like noise. However, some approaches use post-processing to get around the issues brought on by excessive segmentation.

### 1.2 Image Segmentation

#### Overview

Image segmentation approaches search for items that show variations from the surrounding objects or some degree of internal homogeneity. Measures of uniformity and contrast can be used to account for features such as texture, color, and gray level.

<sup>1</sup>Assistant Professor, Department of Computer Science and Engineering (Data Science), Madanapalle Institute of Technology & Science, Madanapalle, Andhra Pradesh – 517325, Email: pugazharasi@gmail.com

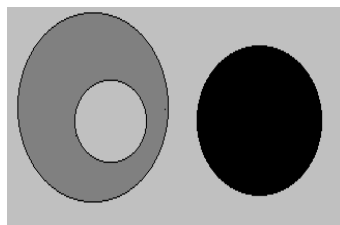
<sup>2</sup>Professor, Department of Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode, Namakkal, Tamilnadu – 637215, Email: k\_sakthivel72@yahoo.com

<sup>3</sup>Associate Professor, Department of Information Technology, Sonal College of Technology, Salem, Tamilnadu-636005, Email: vijimecse@gmail.com

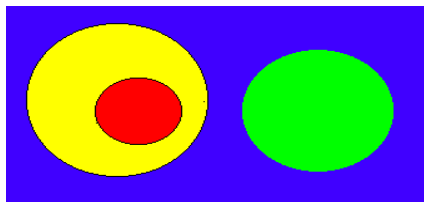
<sup>4</sup>Assistant professor, Department of Artificial Intelligence and Data Science, R P Sarathy Institute of Technology, Salem, Tamilnadu – 636305, Email: pushpaayuvraj@gmail.com

\*Corresponding Author: K. Pugazharasi

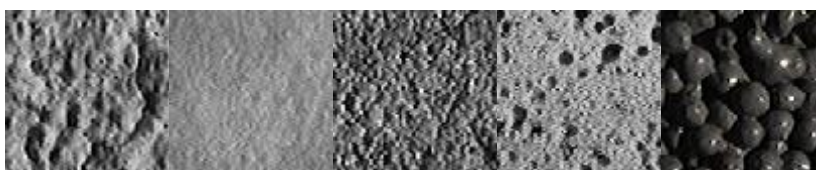
\*Email: pugazharasi@gmail.com



Different gray level



Different color



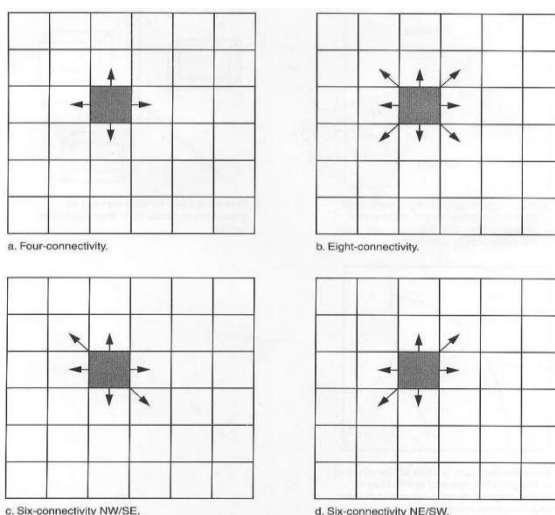
Different Texture

For more segmentation or classification after the first segmentation, one may include higher-level attributes of the object, like perimeter and form. Noise and connectivity issues could have an impact on the segmentation results.

With the previously stated preprocessing technique, noise can be eliminated. Spatial digitalization is the cause of connectivity issues. With proper connectivity

definitions, it can be solved. How can we define what pixels are connected? Connectivity is the term used to describe how one defines a thing. Which pixels in the vicinity are regarded as neighbours must be specified.

A pixel has eight potential neighbours: vertically two, horizontally two and diagonally four. Three definitions of connectedness exist: 1) 4 - connectivity, 2) 8 - connectivity and 3) 6-connectivity.

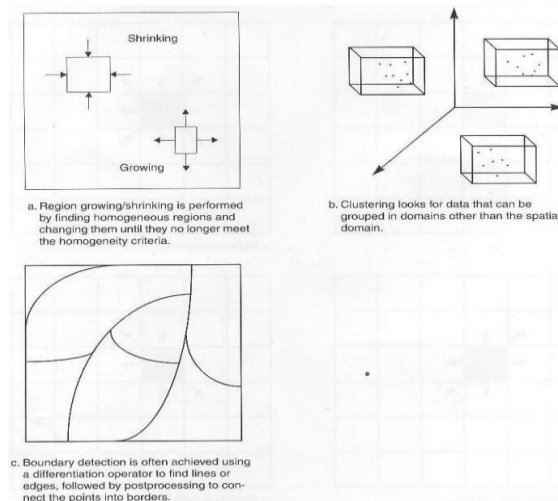


The main categories of image segmentation are as:

1. region growing and shrinking

2. clustering methods

3. boundary detection



While clustering methods can be used in any sector (spatial domain, colour space, feature space, etc.), region growth and shrinking approaches employ a column and row-based space for images (spatial or RC domain). Extensions of line detection methods are boundary detection approaches.

The split-and-merge approach generally works as follows:

1. Specify the homogeneity test. Defining a criterion the area must meet to pass the homogeneity examination entails creating a homogeneity measure that may combine characteristics like brightness, colour, texture, or other application-specific information.
2. Divide the image into areas of similar size.
3. Determine the homogeneity index for each area.
4. If a region's test is successful, an attempt is made to integrate it with its neighbouring regions. The region is divided if the requirement is not satisfied.
5. Keep going until the test is passed in every region.

Starting at a global level, wherein everything in the picture is considered the beginning region, one can proceed in the same manner described before, without any region joining. Alternatively, one may start at the smallest level and only merge without splitting. The result could be quite similar, with the difference in computation time.

However, the overall approach is to look for elements that will be comparable inside an application regardless of the homogeneity test. The grey level variance may be used as the homogeneity metric. The variance essentially serves as a gauge for how wildly different a region's grayscales are. The use of higher-order statistics is possible for features like texture.

### Clustering Techniques

Individual elements in an image are grouped using clustering techniques, which divide the pieces into groups based on a certain amount of internal resemblance.

The primary distinction between these methods and region-expanding techniques is that the principal domain for clustering may be considered a domain different from the RC-based (also known as the spatial domain).

Colour spaces, histogram spaces, and complicated feature spaces are a few of these other domains. By choosing the centre or median along every axis and splitting the area of interest here, the most straightforward clustering approach is dividing the space into sections.

Another method uses a threshold of histograms to segment images. A set of histograms is calculated for a specific set of traits, and each of these histograms is then examined for clear peaks. On the basis of this threshold of the histogram, the best peak is chosen, and the resulting picture is divided into areas.

The first method based on these ideas goes like this:

1. When computing the histograms for every significant element (such as the red, green, and blue for a colour image), treat the whole picture as one region.
2. Run a test for peak location on each histogram. Pick the ideal peak, then place limits on the opposite side of it. Based on this peak, divide the pic into two parts.

3. Smooth the binary limit image until there is just one remaining subregion that is connected.
4. Continue the previous procedures for every region unless no new subregions can be formed or until no histograms contain notable peaks.s:'

## 2. Literature Survey

KNN is an easy-to-use, low-cost computational approach—a system for classifying medical images automatically. The medical image is classified into normal and pathological using a KNN classifier—this notion by R. J. Ramteke and others [1]. Khushboo Singh et al. suggested the SVM classifier-based MRI image classification method [2]. Modern classification methods use support vectors. An algorithm for supervised learning is the support vector machine. In SVM, quadratic programming is used to carry out the classification. Implementation of an effective watershed-based and threshold-based segmentation system for detecting brain tumours was suggested by A . Mustaqeem et al. [3]. Medical imaging techniques were used in this study to find brain tumours. The organic nervous system served as the inspiration for the mathematical problem known as ANN. A survey on the algorithm for detecting brain tumours and their locations in the brain was suggested by Priyanka et al. [4].

Shweta Jain et al. [5] extracted features using the GLCM technique and then categorized them using an ANN. P.Vasuda et al. [6] devised a method to identify tumors in MRIs using fuzzy clustering. Although this approach uses fuzzy C-means, its main flaw is the time needed for computation. Sindhu Shree has created a brain tumour segmentation approach, K.S. et al. [7] and verified on two-dimensional MRI data. Additionally, a three-dimensional image of the discovered tumours is provided. Morphological operations, thresholding, high-fidelity filtering, histogram equalization, linked component labelling, and segmentation were used to find tumors. The total area of the tumor was determined, and the two-dimensional retrieved tumour images were reconstructed into three dimensions. Gopal et al. [8] detected brain tumours from MRIs using algorithms for image processing, including fuzzy, C means and intelligent optimization techniques.

It is meant to identify brain tumors using MRI and computational image clustering techniques such as fluffy C Means and optimization tools like Genetic Algorithm and Particle Swarm Optimization. GA's standard deviation for the classification error of results is 0.078%. GA has an average accuracy of 89.6%. The accuracy in this case is 92.8%, the typical error in the classification of PSO is 0.059%, and the rate of tumour identification is 98.87%. As a result, we noticed that the average

categorization error decreases as the sample size increases. This paper has given convincing evidence that the PSO algorithm worked well for segmenting brain tumours. A new technique based on threshold adaptation was proposed by Badran et al. [9] and can be utilized as a next decision-making tool for surgeons. It detects if an input MRI head image depicts a normal brain or a tumour brain, and based on that determination, it specifies whether the tumour is malignant or benign. Dalal and Triggs introduced HOG descriptors in image classification [10].

A method employs multi-class support vector machine to predict the status of bearing on rolling using vibration signals [11].

Automatic or automatic classification approaches are necessary to distinguish between various tumour types because human observations can result in classification errors [12].

## Research Methodology

### 3.1 Previous Methodology

According to prior research, it is challenging to automatically identify brain tumours in magnetic resonance imaging (MRI) because of the variety and complexity of these lesions' locations, sizes, shapes, and textures. Some methods use tasking-spectral anatomical MRI scans since brain lesions and healthy tissues have similar intensities. A method for detecting tumour tissues using single-spectral anatomic MRI images must be developed, nonetheless, due to the time and financial constraints on obtaining multi-spectral MRI scans and other challenges.

We employed an automatic approach to identify slices with tumours and define the tumour region. The experimental findings on a single contrast mechanism show the effectiveness of our suggested method in brain tumour tissues, which can be accurately segmented with little computing complexity [13]. We also provide a study that compares the performance of statistical features to Hertz wavelet features using a variety of classifiers. Being the first to contrast these assortments of characteristics for tumour segmentation applications, our contribution closes a gap in the literature.

### Texture-based feature extraction techniques

In this work, two feature sets are compared. The Gabor wavelet transform was utilized to produce the Gabor wavelet characteristics, which make up the first set. The GLCM, GLRLM, HOG, and LBP methods—various texture-based feature extraction techniques—extract statistical features for the second collection [14]. In this part, we briefly discuss several texture-based feature extraction methods.

## Statistical feature extraction methods

### First-order statistical features

First-order statistical characteristics such as mean, median, average contrast, energy and entropy, skewness, and kurtosis are helpful. The mean is the picture's average intensity value. The intensity fluctuations around the mean are shown by variance. The asymmetry of the histogram around the mean is measured by skewness. The histogram's flatness is known as kurtosis. The unpredictability of intensity values is revealed by entropy.

### Gray-level co-occurrence matrix features

The association between pixels or groups of pixels (often two) is reflected by spatial grey-level co-occurrence, which calculates image attributes related to second-order statistics. The appearance of pairs of photons that separate by a specific distance,  $d$ , is described by the GLCM, a 2D histogram. Let  $x_1; y_1$  and  $x_2; y_2$  be two separate pixels with grey level intensities  $i$  and  $j$ , respectively, and let  $I; y$  be an image of size  $NM$ ,  $G_{gray}$  levels, and grey levels.

### Grey-level run length method features

A geographic domain grey level value is given a quantitative parameter via the domain of space second-order statistical procedure known as GLRLM. The most extensive collinear-associated set of dots with a single grey level is what is referred to in GLRLM as a texture primitive termed the grey level run length. The length and trajectory of the loop for a specific grey value define the grey level runs.

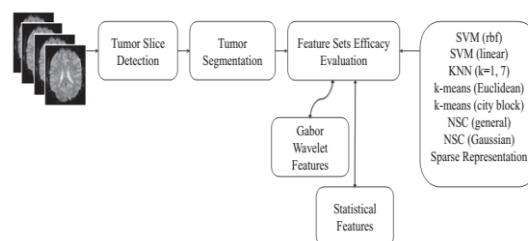
### Oriented gradient features: Histogram

Features of HOG are descriptors that are extensively employed in the processing of images and computer vision for object detection. These descriptors are justified by the idea that the arrangement of luminosity gradients or edge orientations can adequately characterize the appearance and shape of nearby objects. The method tallies instances of orientations with gradients in particular areas of a picture [25]. HOG employs overlapping local contrast normalization for greater precision on a dense grid of cells with equal spacing.

In HOG, a picture is partitioned into discrete, connected areas known as cells. Then, a histogram of gradient or edge orientations for each cell's individual pixels is compiled for each cell. The description is then created by combining these histograms. By computing an indicator of the intensity across a broader area, known as a block, and using the result to normalize all cells within the block, the surrounding histograms can be compare-normalized for increased accuracy. This normalization provides lower sensitivity to variations in light or shadowing. In this work, 80 HOG features are derived.

### Linear binary pattern features

By thresholding the neighbourhood of the window's central pixel with a single value and setting binary values for its neighbours, the LBP operator sweeps the window over the picture and assigns labels to its central pixel. The LBP then computes the sum of the powers of two multiplied by the binary values, increasing either clockwise or anticlockwise. As a texture description, a graph of these 256 unique labels is employed. There are several sizes that might be considered neighbourhoods. Any range and any quantity of nearby pixels may be used.

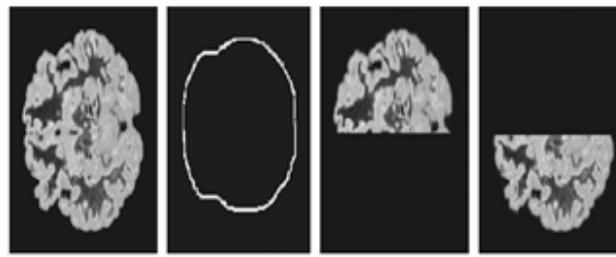


**Fig 3.1** Block Diagram of the Methodology

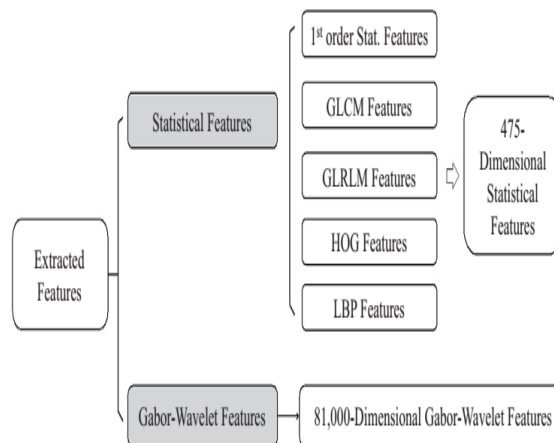
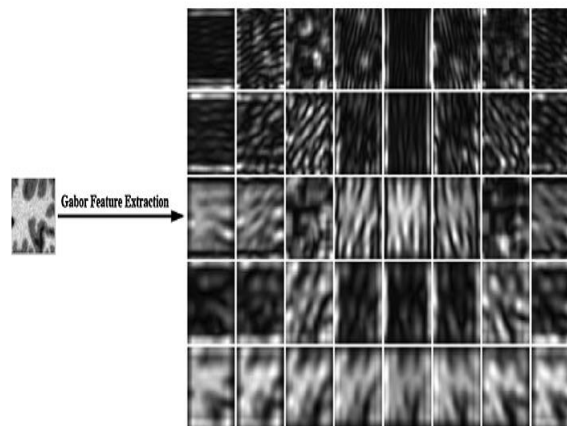
The final image is significantly influenced by the wavelet family. Different wavelet families have distinctive qualities that support various attributes in the combined image. The degree of decomposition that will be used is another crucial element because as the amount

of decomposition increases, features are lost or the degree of reconstruction changes. The best combination of the aforementioned qualities has

been carried out experimentally, and the same has been discussed in a later section.



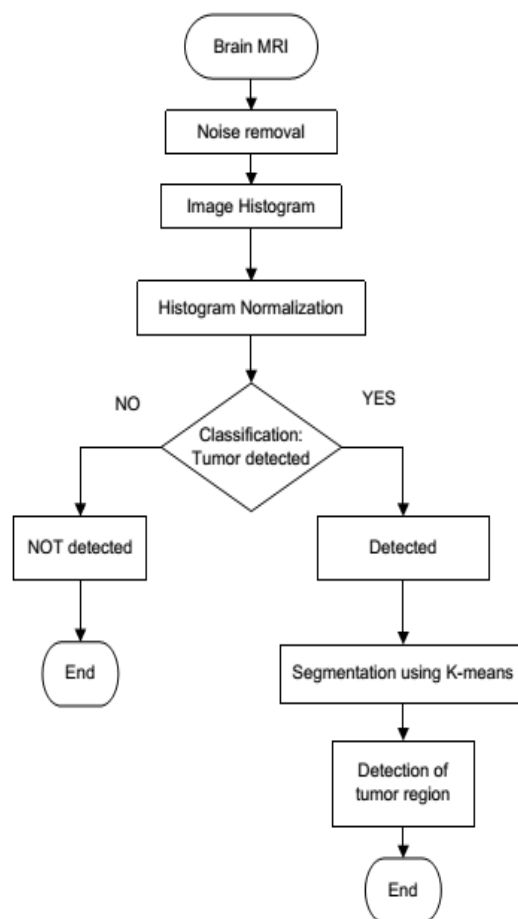
**Fig (i)** Original image, (ii) Projects the brain borderline, (c) and (d) Brain hemispheres examples.



### 3.2 Proposed Methodology

This research applies a unique technique that combines K-means segmentation and histogram normalization. The supplied image is first pre-processed to get rid of any undesirable signals or noise. The pre-processed image's

histogram is normalized, and MRI classification is carried out. The image is segmented using the K-means method to remove the tumour from the MRI. Utilized for implementation is MATLAB. Processing the data is crucial to eliminate the unwanted MRI components before using the technique for the image.



### A. Median Filter

The median filter is composed of digital and non-linear filters. It removes any unwanted noise or signals from the image. The values of every pixel in the neighbourhood are arranged in descending order, from low to high. Next, the median is calculated. The median filtering technique is advantageous when taking into account pixels close to the margins since it preserves the edges and therefore prevents the generation of fake pixel values. When dealing with impulse noises, the median filter works the best.

### B-type adaptive filter

An adaptive filter is a linear filter. It improves images with abrupt brightness shifts and eliminates impulsive and speckle noise.

### C. Averaging filter

One of these most straightforward filters for image smoothing is the averaging filter. It is a low-pass linear filter. The parameters of the pixels that are not beneficial to the portrayal of their surroundings are removed by this filter. It operates as a kernel-based convolution filter.

The softening of the image increases as the kernel size increases.

### D. Gaussian filter

This filter, also known as a smoothing operator, is used to distort images. It eliminates the minute features that are built into the image. A Gaussian function that describes the likelihood distributions of the noise is its impulse response. Gaussian noise can be eliminated using it.

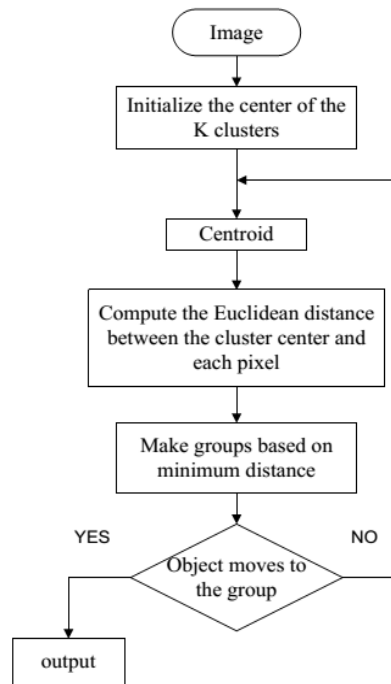
### E. Un-sharp masking filter

In essence, the un-sharp masking filter is a sharpening operator that is applied to the input image to improve the edges by deleting the unsharp version of the picture.

## CLASSIFICATION

### *K-MEANS SEGMENTATION*

The practice of breaking an image into several parts or segments is called picture segmentation. The aim of this strategy is to convert a picture into a more easily researched, meaningful shape. K-means and other unsupervised segmentation techniques are employed. It creates k clusters out of a group of information signals.



K-means is one of the most straightforward unsupervised learning techniques for resolving the well-known clustering problem. A simple and direct method employing a preset number of clusters (let's assume  $k$  clusters) is used to classify the provided data set. The primary idea is to build  $k$  centroids, one for each cluster. Since different sites offer different impacts, it is essential to place these centroids carefully. The best course of action is to put them as far apart as possible.

Next, every point extracted from a specific data set is joined to the nearest centroid. When no points are remaining available, the initial stage and young age are concluded. The task is to select new centroids that will act as the barycenters of the clusters created from the points in the same data set closest to the nearest new centroid. A loop has been established. As we may see, this loop causes the  $k$  centroids to progressively change locations until no further modifications are performed. In other words, centroids have stopped moving. Finally, and perhaps most importantly, the algorithm's goal is to minimize a subjective function—in this case, the squared error ratio.

#### Advantages:

- In this case, a single image frame is used to map the benefits of the wavelet transform and filter.
- The suggested method improves fusion quality by fusing anatomical and functional information with multi-level decomposition.
- For an enormous amount of database pictures, we completed MR-PET, MRSPECT, MR T1-T2 and MR-CT modal fusion here.

The optimum scale values are selected using Optimization.

### 4. Algorithms and System Design

#### 4.1 Algorithms

##### 4.1.1 FILTERING

##### Median Filter Pseudo Code

Code for a simple 2D median filter algorithm might look like this:

```

allocate outputPixelValue[image width][image height]
allocate window>window width * window height
edgex := (window width / 2) rounded down
edgey := (window height / 2) rounded down
for x from edgex to image width - edgex
for y from edgey to image height - edgey
    i = 0
    for fx from 0 to window width
    for fy from 0 to window height
        window[i] := inputPixelValue[x + fx - edgex][y + fy - edgey]
    i := i + 1
    sort entries in window[]
    outputPixelValue[x][y] := window>window width * window height / 2]
  
```



## Gaussian Filter

The suggested method improves fusion quality by fusing anatomical and functional information with multi-level decomposition. For an enormous amount of database pictures, we completed MR-PET, MRSPECT, MR T1-T2 and MR-CT modal fusion here.

The Gaussian Blur method is one of numerous blur implementation algorithms. To process images, Gaussian distribution is used. This article will teach you about the

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

Here,  $\mu$  is the average of  $x$ ,  $\mu$  equals 0.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}$$

Based on this, the 2D-Gaussian function is derived as

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

This function can be used to calculate each point's weight.

### 2.1.1 NORMALIZED HISTOGRAM

#### Transformation of Histogram

##### A: Normalization of a Histogram

The suggested method improves fusion quality by fusing anatomical and functional information with multi-level

$$\begin{aligned} \{[0,L-1] \rightarrow N_x \rightarrow \text{Card}(x) \text{ becomes} \\ \{[0,L-1] \rightarrow [0,1] x \rightarrow \text{pdf}(x) = \text{Card}(x) \\ \sum_{i=0}^{L-1} \text{Card}(x_i) \{[0,L-1] \rightarrow N_x \rightarrow \\ \text{Card}(x) \text{ becomes } \{[0,L-1] \rightarrow [0,1] x \rightarrow \\ \text{pdf}(x) = \text{Card}(x) / \sum_{i=0}^{L-1} \text{Card}(x_i) \end{aligned}$$

where *Card* is the cardinality of the set. Here, it represents the number of pixels.

#### Pseudo Code

```
update_hist(curr_population_size, hist,
new_element) {
    index = determine_index(new_element);
    new_pop_size = curr_pop_size + 1;
    for i = 1 to number_of_bins(hist)
        if i == index then
            hist[i] = (hist[i]*curr_pop_size + 1) /
            new_pop_size;
        else
            hist[i] =
            (hist[i]*curr_pop_size)/new_pop_size;
        endif
    return hist
}
```

Now Suppose, hist[1] = 0.3, hist[2] = 0.2, and hist[3] = 0.5. Then, by the above pseudo-code, a new value, when

Gaussian Blur method, a straightforward algorithm. It is a form of data smoothing that may be applied in various circumstances.

## Gaussian function

A two-dimensional normal distribution is required. The Gaussian function is the name given to the density function of a normal distribution. The format with one dimension is:

decomposition. For an enormous amount of database pictures, we completed MR-PET, MRSPECT, MR T1-T2 and MR-CT modal fusion here.

A histogram is normalized when the discrete variance in strengths is transformed into a discrete range of probabilities. We must divide each histogram value by the total number of pixels in order to do this. Digital images are a matrix since they are separate sets of data. Dividing each  $n_k$  by the size of the array—the total width and length of the image—is equivalent to doing this.

$$\begin{aligned} n_k &= n_{\text{length}} \times \text{width} \\ &= \text{pr}(r_k) n_k \\ &= n_{\text{length}} \times \text{width} \\ &= \text{pr}(r_k) \end{aligned}$$

The above can be represented mathematically as follows:

it arrives and whose bin is determined to be, say 2, updates the hist as:

$$\text{hist}[1]=(0.3*100)/101= 0.3069306930693069$$

$$\text{hist}[2]=(0.2*100+1)/1010.2079207920792079$$

$$\text{hist}[3]=(0.5*100)/101=0.4950495049504950$$

A word of caution: Rounding problems generated by floating-point arithmetic may cause repeated divisions and divisions with big integers to produce incorrect values.

## 2.1.2 K MEANS Algorithm

### PSEUDO CODE FOR IMPROVED K MEANS

---

**Algorithm 1: K-means Algorithm**

---

**Data:**  $N$  Data points  $d_1, d_2, d_3, \dots, d_N$

**Result:**  $N$  Data points in  $K$  clusters

- 1 Randomly Initialize  $K$  cluster centers  $\mu_1, \mu_2, \mu_3, \dots, \mu_k;$
- 2 **while** *Not convergence (no changes in the clusters or maximum iteration number)* **do**
- 3     Assign each point  $d_i$  to the cluster with nearest  $\mu_j;$
- 4     Calculate the new mean  

$$\mu_j = \frac{1}{|C_j|} \sum_{d_i \in C_j} d_i;$$
- 5 **end**

---

The improved algorithm works on tiny samples compared with the whole dataset and requires much less repetition. Early sampling costs are insignificant in the overall method for massive datasets. The time spent clustering the entire dataset using the initial centres that were found makes up most of the upgraded K-means' computational cost.

$K$ \*clusters  $(O(nd))$ , and the time merging  $K$ 'clusters into  $K$  clusters  $O(nd(k' - k))$ . So, the computational complexity of the improved algorithm is  $O(ndk')$ .

## 5. Results and Discussion

An outstanding performance language for scientific computing is called MATLAB. In a simple-to-use environment, it mixes computing, visualization, and programming while expressing issues and solutions using well-known mathematical notation.

The recommended research approaches are evaluated in this part using the MATLAB environment to estimate their ability to detect faults.

To determine the performance measure values for each work, it is implemented and evaluated under various setup conditions. The suggested research methodology is

the Normalisation of Bitmap and K-means algorithm Segmentation Procedure (NH\_KM). The proposed research methodology is contrasted with previously published work, specifically Gabor wavelet (GW) domains and C-means.

The performance metrics "Accuracy, sensitivity, and specificity" are taken into account while assessing how well the suggested research approaches are doing. The following subsections provide explanations and illustrations of the performance metrics comparison results.

### Accuracy (%)

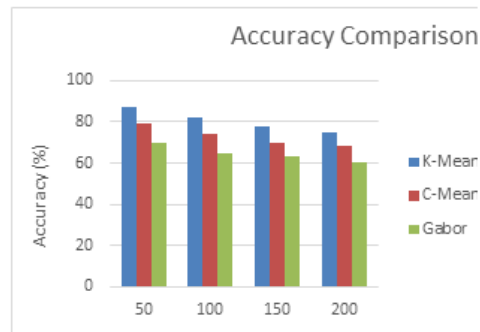
When all the real classification parameters ( $T_p + T_n$ ) are divided by the sum of the classification parameters ( $T_p + T_n + F_p + F_n$ ), the accuracy of the model is defined as its overall correctness. The accuracy is calculated in this manner:

$$\text{Accuracy} = \frac{T_p + T_n}{(T_p + T_n + F_p + F_n)}$$

Where  $T_p$  - True positive

$T_n$ -True negative

$F_n$ -False negative



**Fig 1** Accuracy Comparison

No of input Samples	Accuracy (%)		
	K-Means	C-Means	Gabor
50	87	79	70
100	82	74	65
150	78	70	63
200	75	68	60

**Table 1.** Accuracy Measure

Figure 1 above shows how the comparison metric's correctness is assessed using both the existing and suggested methods. The algorithms are used for the x-axis, and their precision value is represented on the y-axis. For the specified transaction database, the suggested K-Means approach outperforms the current C-Means and Gestalt methods in terms of accuracy. The outcome demonstrates that the proposed technique uses K-Means to improve utility results.

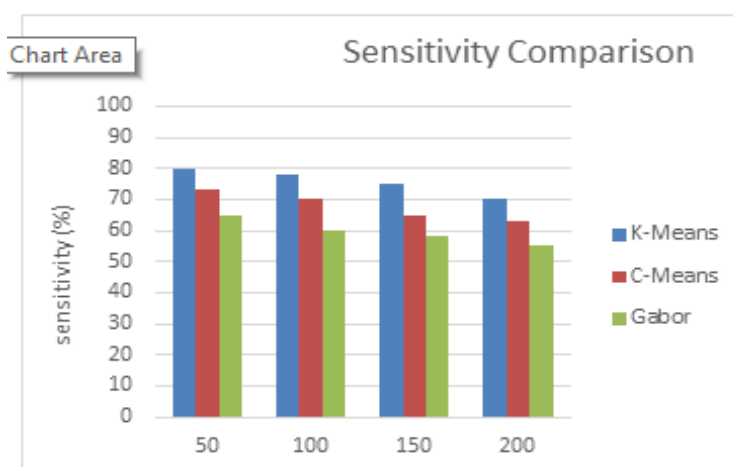
**Sensitivity (%)**

Sensitivity is the number of positivity accurately classified as such (for example, the proportion of ill

patients correctly classified as having the illness)[15]. It is also known as the real positive rate, recollection, or the rate of detection in various fields. Sensitivity is the test's capacity to accurately identify classifiers that can identify the presence of bugs. This can be written mathematically as:

$$\text{Sensitivity} = \frac{\text{Number of true positives}}{\text{Number of true positives} + \text{Number of false negatives}}$$

The graphical representation of the sensitivity measurement values of the proposed research methodology is given in Figure 2.



**Fig 2** Sensitivity Comparison

No of input Samples	Sensitivity (%)		
	K-Means	C-Means	Gabor
50	80	73	65
100	78	70	60
150	75	65	58
200	70	63	55

**Table 2.** Sensitivity Measure

Sensitivity measure evaluations of the suggested research approaches are shown in Figure 2. This graph demonstrates how the proposed research methodology can efficiently and with increased performance forecast the flaws in the software. Compared to C-Means and Gabor, the suggested method, K-Means, produces better results.

### Specificity (%)

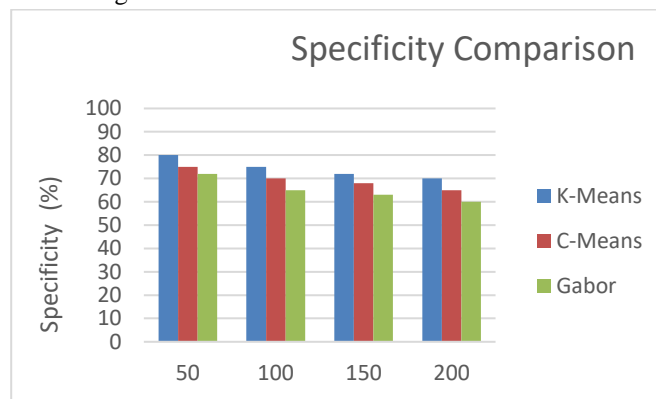
Specificity, sometimes called the true negative rate, quantifies the percentage of negatives correctly categorized as such (i.e., the percentage of healthy persons correctly classified as having none of the

disease). The ability of the test to recognize a classifier with a condition is known as specificity. This may be stated mathematically as follows:

Specificity

$$= \frac{\text{number of true negative}}{\text{number of true negatives} + \text{number of false positives}}$$

The graphical representation of the specificity measurement values of the proposed research methodology is given in Figure 3.



**Fig 3** Specificity Comparison

No. of input Samples	Specificity (%)		
	K-Means	C-Means	Gabor
50	80	75	72
100	75	70	65
150	72	68	63
200	70	65	60

**Table 3.** Specificity Measure

In Figure 3, specificity measure comparisons of the proposed research methodologies are given. This graph proves that the proposed research method can efficiently predict the faults present in the software with improved performance. This comparison analysis shows that the

proposed method, K-Means, has better outcomes than C-Means and Gabor.

### 6. Conclusion

Given that brain tumours are one of the most well-known brain illnesses, it is crucial for medical analysis to

identify and classify the tumour. The k-clustering algorithm is used to segment the brain tumour from the MRI.

When calculating MSE, the averaging filter yields the most significant result (MSE=0.0038), while the median filter works best for noise reduction (PSNR= 78.7316). Based on MSE, a comparison is conducted between the Median, Adaptive, Averaging, Gaussian, and Un-sharp mask filters. After histogram normalisation, the images were classified as "tumor image" and "non-tumor image" using SVM and the Naïve Bayes classifier. SVM efficiency is 91.49%, while Naïve Bayes efficiency is 87.23%. It is determined that SVM outperforms the Naive Bayes classifier in terms of efficiency.

The suggested method has some drawbacks because, in some tumour photos, the outcomes could have been better, and the tumour could not be accurately detected. The exact or correct boundary of the tumour zone could not be determined by the algorithm. The programme has room for improvement to more accurately detect tumours. The limitations of the proposed method can be overcome in the future, and the calibre of the images generated can be improved by utilizing superior morphological techniques.

#### References

- [1] E. El-Dahshan, H. Mohsen, K. Revett and A. Salem, "Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm", *Expert Systems with Applications*, vol. 41, no. 11, pp. 5526-5545, 2014.
- [2] N. Nabizadeh and M. Kubat, "Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features", *Computers & Electrical Engineering*, vol. 45, pp. 286-301, 2015.
- [3] R. Gonzalez, R. Woods and B. Masters, "Digital Image Processing, Third Edition", *J. Biomed. Opt.*, vol. 14, no. 2, p. 029901, 2009.
- [4] K. Mardia and T. Hainsworth, "A spatial thresholding method for image segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 6, pp. 919-927, 1988.
- [5] K. Qureshi and S. Hussain, "A Comparative Study of Parallelization Strategies for Fractal Image Compression on a Cluster of Workstations", *Int. J. Comput. Methods*, vol. 05, no. 03, pp. 463-482, 2008.
- [6] M. Wu, C. Lin and C. Chang, "Novel image copy detection with rotating tolerance", *Journal of Systems and Software*, vol. 80, no. 7, pp. 1057-1069, 2007.
- [7] H. Rajini.N and B. Bhavani.R, "Automatic MR Brain Tumor Detection using Possibilistic C-Means and K-Means Clustering with Color Segmentation", *International Journal of Computer Applications*, vol. 56, no. 2, pp. 11-17, 2012.
- [8] B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation", *J. Electron. Imaging*, vol. 13, no. 1, p. 146, 2004.
- [9] M. Waseem Khan, "A Survey: Image Segmentation Techniques", *IJET*, pp. 89-93, 2014.
- [10] S. Madhukumar and N. Santhiyakumari, "Evaluation of k-Means and fuzzy C-means segmentation on MR images of brain", *The Egyptian Journal of Radiology and Nuclear Medicine*, Vol. 46, no. 2, pp. 475-479, 2015.
- [11] Rajeswari C., Sathiyabhama B., Devendiran S., Manivannan K. Bearing fault diagnosis using wavelet packet transform, hybrid PSO and support vector machine, *Procedia Engineering*, Vol.97(1) PP:1772-1783, 2014.
- [12] Rajeswari C., Sathiyabhama B., Devendiran S., Manivannan K. A Gear fault identification using wavelet transform, rough set based GA, ANN and C4.5 algorithm. *Procedia Engineering*, Vol – 2 PP: 338-344 , DOI : 10.1016/j.procs.2010.11.044,2014
- [13] Sudha S., Suresh G.R., Sukanesh R. Speckle noise reduction in ultrasound images using context-based adaptive wavelet thresholding, *IETE Journal of Research*, Vol-55 (3) PP135-143, DOI: 10.4103/0377-2063.54905 2009.
- [14] Rampriya R., Suguna N., Kumar R. G. S, & Sudhakar P (2023). Optimal Feature Selection and Classification of Respiratory Diseases by Novel EFICNN-EBOA Algorithm: A Real Time Implementation Concept, *International Journal of Intelligent Systems and Applications in Engineering*, 12(2s), 699–712.
- [15] Shanthy T., Sabeenian R.S. Modified Alexnet architecture for classification of diabetic retinopathy images, *Computers and Electrical Engineering*, Vol-76 PP 56-64, DOI:10.1016/j.compeleceng.2019.03.004