j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 5, No:1. 2022



# CLUSTERING OF MEDICAL IMAGES USING MULTIWAVELET TRANSFORM AND K-MEANS ALGORITHMS

Waleed A. Mahmoud Al-Jawher*[1], Sh. A. Shaban[2]

*[1]College of Engineering, Uruk University, Baghdad, Iraq*
*[2] Informatics Institute for Postgraduate Studies, Baghdad, Iraq.*

*Profwaleed54@gmail.com*

**Abstract** Imaging data and biological records are both included in the field of medical informatics. Medical image clustering is an important area of research that is getting increased interest in both academia and the health professions. It addresses the issues of medical diagnosis, analysis, and education. Several medical imaging methods and applications based on data mining have been made and tested to deal with these problems. This paper looks at how image classification techniques diagnose diseases in the human body. It concerns the imaging modalities, dataset, and the pros and cons of each method. An optimal clustering technique of medical images using multiwavelet transforms proposed that combines the multiwavelet transform filterbanks with the k-means clustering algorithm to improve the performance and get a clinically meaningful clustering shape. In comparison with other methods of clustering, it was shown that this method has a much higher cluster classification than those published before. A user-friendly Matlab program has been constructed to test and get the results of the proposed algorithms

Crossref 10.36371/port.2022.1.5

**Keywords:** *Clustering, Medical data, Feature extraction, K-means clustering, Multiwavelet algorithm.*

## 1. INTRODUCTION

Medical classification of images is an exciting area of study because it combines the problems of diagnosis and analysis of medical images [1]. Data clustering is the process of grouping together sets of related data. A clustering method divides a data set into groups so that group similarities are more significant than group differences. [2]. These bits of information can be called data points, observations, or feature vectors. The data in each cluster are similar to each other but different from the data in other clusters. Clustering can be used to find information, compress data, study the weather, physiology, and medicine, and even run a business. Therefore, image clustering is the process of putting together groups of images from a database. After grouping, each picture in the database has a "class label" that tells us how similar this image is to other images with the same class label. The main goal of image clustering is to divide image databases or archives into groups or clusters from which knowledge or predictions can be derived. Image clustering also summarizes and displays the content of images. Moreover, it can be used for unsupervised clustering of an image set or database, image categorization, picture segmentation, content-based image retrieval, and image categorization [3]. In medical image datasets, unsupervised clustering is needed to organize and store many medical images quickly and easily. This makes it easier to classify medical images based on their content [3,4]. In addition, clustering is the most important part of data analysis and mining. As data sets grow and their properties and relationships change, finding groups of objects that are highly related to each other is very useful even when there are a lot of them [5]. A clustering method can be either clear or unclear. A hard clustering algorithm puts each pattern in only one cluster and gives the result during run time. A fuzzy clustering method provides each input with a membership degree based on how close it is to more than one cluster [6,7]. The aim is to present a new way to solve the problem of separating parts of a medical image by using improved clustering techniques based on image processing and data clustering techniques that already exist.

## 2. THE PROPOSED METHOD OF MEDICAL IMAGES CLUSTERING

The block diagram of the proposed medical image clustering algorithm is shown in Fig. (1)


***Fig. (1)** The Block Diagram of Clustering Algorithm*



Waleed A. M. Al-Jawher, Sh. A. Shaban. (2022). Clustering Of Medical Images Using Multiwavelet Transform And K-Means Algorithms. Journal Port Science Research, 5(1), 35–42. https://doi.org/10.36371/port.2022.1.5

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 5, No:1. 2022

The block diagram consists of five stages. The input image is taken from the device. Next, it is treated by several preprocesses and adapted in size and converted into grayscale. Then, feature extraction is performed. In this work, multiwavelet transform is applied to extract useful features. Finally, a clustering algorithm is implemented to work on the multiwavelet coefficients to obtain the required clusters.

A detailed description of the different stages will be given next sections.

## 3. IMAGE ACQUISITION

During the phase of image acquisition, datasets from the UCI machine learning repository are accessed. There will include data sets on Wisconsin breast cancer (wdbc and wpdc), the heart, diabetes, the thyroid, lung cancer, and the initial tumor. After that, the data will undergo procedures that make it ready for usage [9].

Biomedical devices that use imaging techniques such as computed tomography (CT), magnetic resonance imaging (MRI), and mammography provide medical image data. Several medical imaging modalities use ionizing radiation, nuclear medicine, magnetic resonance, ultrasound, and optical techniques as their modality media. Each modality has a unique quality and specific effect on the human body's structure and organs. [1]

There are four imaging modalities:

### 3.1.1. Projectional Imaging.

X-rays are electromagnetic radiation (EM) wavelengths ranging from 0.1 to 10 nm. They are changed into photons with 12–125 keV of energy. The x-ray image processing projection is used almost when lab tests become important to diagnose medical problems. Making an image comprises three main steps: pre-reading the image, reading the main image, and processing the image [1].

### 3.1.2. Computed Tomography (CT).

Minor variations in attenuation (less than 5%) can prevent the typical x-ray imaging projection from producing acceptable results. CT enhances subject to contrast with 1% or less discrimination. CT is commonly used in cancer screening applications, including virtual colonoscopy and lung screening. CT imaging has different variations: PET/CT, CT perfusion, CT angiography, dual-source and dual-energy CT, and PET/CT positron emission tomography [1].

### 3.1.3. Magnetic Resonance (MR).

In the Magnetic Resonance Imaging (MR) technique, a strong magnetic field is used to align the hydrogen atoms in water molecules' nuclear magnetization. Cross-sectional imaging techniques using MR have become the norm for observing both soft tissues (such as muscle, brain, and fat) and hard structures (like bones, especially marrow bone) [1].

### 3.1.3. Ultrasound Imaging.

Cross-sectional images of the human body can be made using high-frequency sound waves between 1 and 20 MHz. The type of tissue the sound waves are passing through determines the strength of the echo ultrasonography. [1].

## 4. DATA PRE-PROCESSING

Data pre-processing is a set of steps taken on raw data to get the best results for a single dataset. It has a significant effect on how well classification algorithms work. So, in the pre-processing phase of this algorithm, the min-max normalization method is used to ensure that the data range is between [1,0]. Normalization is an important part of data preprocessing because some machine learning methods can't directly deal with continuous attributes. First, the change of data in the interval [1,0] is better for humans to understand cognitively. Second, the process of computing will go faster [9]. Unfortunately, there are a few difficulties with image data, such as complexity, accuracy, and sufficient. To get the intended outcomes, the data must be preprocessed (cleaned and processed to the proper format) before creating a computer vision model. We will examine image data preparation in this section, which transforms image data into a format that machine learning algorithms can understand. This is usually used to improve the model's accuracy and simplify it. To preprocess image data, many techniques are used. Examples include image resizing, grayscale image conversion, and image enhancement [9].

## 5. FEATURE EXTRACTION

In this part, the Multiwavelet Transformation algorithm is used. After pre-processing the image, in this case, Multiwavelet decomposition is applied. The Multiwavelet decomposition process of matrix X is discussed briefly below.

### 5. 1. Fundamental of Multiwavelets

n particular, multiwavelets have two or more scaling and wavelet functions in contrast to wavelets, which only have one scaling function $\Phi(t)$ and one wavelet function $\Psi(t)$. For notational convenience, the set of scaling functions can be written using the vector notation $\phi(t) = [\phi_1(t),\ \phi_2(t) \dots \phi_r(t)]^T$ Where $\Phi(t)$ is called the multiscaling function. Likewise, the multiwavelet function is defined from the set of wavelet functions as $\psi(t) =$



Waleed A. M. Al-Jawher, Sh. A. Shaban. (2022). Clustering Of Medical Images Using Multiwavelet Transform And K-Means Algorithms. Journal Port Science Research, 5(1), 35–42. https://doi.org/10.36371/port.2022.1.5

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS
Journal port Science Research
Available online www.jport.co
Volume 5, No:1. 2022

$[\psi_1(t), \psi_2(t) \ldots \psi_r(t)]^T$. When r=1, ψ(t) is called a scalar wavelet or simply a wavelet. While in principle, r can be arbitrarily large. The majority of multiwavelets researched up to this point are for r=2 [9]. Similar to scalar wavelets, the multiwavelet two-scale equations

$$\phi(t) = \sqrt{2}\sum_{K=-\infty}^{\infty} H_K \phi(2t - K) \quad \ldots (1)$$

$$\psi(t) = \sqrt{2}\sum_{K=-\infty}^{\infty} G_K \psi(2t - K) \quad \ldots (2)$$

Note, however, that $\{H_K\}$ and $\{G_K\}$ are matrix filters, i.e., $H_K$ and $G_K$ are r × r

$$\begin{bmatrix}\phi_1(t)\\ \phi_2(t)\end{bmatrix} = \sqrt{2}\sum_K H_K \begin{bmatrix}\phi_1(2t-K)\\ \phi_2(2t-K)\end{bmatrix} \quad \ldots (3)$$

$$\begin{bmatrix}\psi_1(t)\\ \psi_2(t)\end{bmatrix} = \sqrt{2}\sum_K G_K \begin{bmatrix}\psi_1(2t-K)\\ \psi_2(2t-K)\end{bmatrix} \quad \ldots (4)$$

where $H_K$ for the GHM system are four scaling matrices $H_0$, $H_1$, $H_2$, and $H_3$, [12],

$$H_0 = \begin{bmatrix}\frac{3}{5\sqrt{2}} & \frac{4}{5}\\ -\frac{1}{20} & -\frac{3}{10\sqrt{2}}\end{bmatrix},\ H_1 = \begin{bmatrix}\frac{3}{5\sqrt{2}} & 0\\ \frac{9}{20} & \frac{1}{\sqrt{2}}\end{bmatrix},\ H_2 = \begin{bmatrix}0 & 0\\ \frac{9}{20} & -\frac{3}{10\sqrt{2}}\end{bmatrix},$$

$$H_3 = \begin{bmatrix}0 & 0\\ -\frac{1}{20} & 0\end{bmatrix}$$ …(5), where $G_K$ for the GHM system are four scaling matrices $G_0$, $G_1$, $G_2$, and $G_3$, [12], $G_0 =$

$$\begin{bmatrix}-\frac{1}{20} & -\frac{3}{10\sqrt{2}}\\ \frac{1}{10\sqrt{2}} & \frac{3}{10}\end{bmatrix},\ G_1 = \begin{bmatrix}\frac{9}{20} & -\frac{1}{\sqrt{2}}\\ -\frac{9}{10\sqrt{2}} & 0\end{bmatrix},\ G_2 = \begin{bmatrix}\frac{9}{20} & -\frac{3}{10\sqrt{2}}\\ \frac{9}{10\sqrt{2}} & -\frac{3}{10}\end{bmatrix},$$

$$G_3 = \begin{bmatrix}-\frac{1}{20} & 0\\ -\frac{1}{10\sqrt{2}} & 0\end{bmatrix} \quad \ldots(6)$$

### 5.2 Preprocessing for Multiwavelets

A tree-structured matrix filter bank that operates on vector sequence inputs instead of scalar ones can conduct multiwavelet transformation (lower resolution coefficients can be derived from higher resolution coefficients by a tree-structured technique called a filter bank [17]). Unlike scalar wavelet systems, then, preprocessing is often necessary to extract vector sequence input from the signal to improve performance [15]. Numerous techniques exist for producing such a signal from 2-D signal data[10]. Oversampled approach, a 2-D approximation-based preprocessing scheme for GHM discrete multiwavelet transform of 2-D signals, is developed in this paper called an Oversampled Scheme

### 5.3. Oversampled Scheme: Tretreatment of Repeated Rows

In multiwavelet situations, GHM multiscaling and multiwavelet function coefficients are 2×2 matrices that must multiply vectors throughout the transformation stage (instead of scalars). Therefore, a multifilter bank requires two input rows. Thus, signal repetition is the most obvious approach to obtaining two input rows from a given signal. There are two rows in the multifilter bank. This approach, known as "Repeated Row," introduces a two-factor oversampling of the data. To perform repeated row preprocessing on a scalar input signal {X_K } of length N (where N is supposed to be a power of 2 and is, therefore, an even number), we need to replicate the input stream with the same stream multiplied by a constant α. Thus, the input length-2 vector is formed from the original as,

$$\begin{bmatrix}X_K\\ \alpha X_K\end{bmatrix} \quad where\ K = 0, 1, 2, \ldots, N-1 \qquad \ldots (7)$$

Here $\alpha$ is a constant; it is commonly set so that if $X_K = C = constant$, the output of the high-pass multifilter is zero for every k. This is always possible if the system's approximation order is greater than zero. For the GHM case, $\alpha = 1/\sqrt{2}$ is selected since:

$$[G_0 + G_1 + G_2 + G_3]\begin{bmatrix}C\\ \frac{C}{\sqrt{2}}\end{bmatrix} = \frac{1}{10}\begin{bmatrix}8 & -\frac{10}{\sqrt{2}}\\ 0 & 0\end{bmatrix}\begin{bmatrix}C\\ \frac{C}{\sqrt{2}}\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix} .. (8)$$

The output from the low-pass multifilter is simply a scaled version of the input,

$$[H_0 + H_1 + H_2 + H_3]\begin{bmatrix}C\\ \frac{C}{\sqrt{2}}\end{bmatrix} = \frac{1}{5}\begin{bmatrix}\frac{6}{\sqrt{2}} & 4\\ 4 & \sqrt{2}\end{bmatrix}\begin{bmatrix}C\\ \frac{C}{\sqrt{2}}\end{bmatrix} = \sqrt{2}\begin{bmatrix}C\\ \frac{C}{\sqrt{2}}\end{bmatrix} .. (9)$$

Following the multiwavelet reconstruction (synthesis) step is the post-filtering process.

### 5.4 General Formula for Calculating DMWT Using an Oversampled. Preprocessing Scheme (Repeated Row Preprocessing)

By utilizing an oversampled preprocessing approach (repeated row preprocessing), the DMWT (discrete multiwavelet) matrix is doubled in size relative to the input, which should be a square matrix $N \times N$ where N is a power of 2. After preprocessing, transformation matrix dimensions equal image dimensions. To compute a 2-D Discrete Multiwavelet Transform with a single level, the following steps must be taken:

**5.4.1. Verifying input signal dimensions**: The input matrix must be square, $N \times N$, with N being a power of 2. Therefore, the initial phase of the transform technique is to verify the dimensions of the input matrix. If the input is not a square matrix, the rows or columns of zeros must be added to create a square matrix.

**5.4.2. THE BUILDING OF A TRANSFORMATION MATRIX**:


Waleed A. M. Al-Jawher, Sh. A. Shaban. (2022). Clustering Of Medical Images Using Multiwavelet Transform And K-Means Algorithms. Journal Port Science Research, 5(1), 35–42. https://doi.org/10.36371/port.2022.1.5

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 5, No:1. 2022

For computing Discrete Multiwavelet Transform, using the following transformation matrix

$$\begin{bmatrix} H_0 & H_1 & H_2 & H_3 & 0 & 0 & \dots \\ G_0 & G_1 & G_2 & G_3 & 0 & 0 & \dots \\ 0 & 0 & H_0 & H_1 & H_2 & H_3 & \dots \\ 0 & 0 & G_0 & G_1 & G_2 & G_3 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} \quad \dots (10)$$

Where Hi and Gi are the low- and high-pass filter impulse responses. The GHM low- and high-pass filter matrices described in (5) and (6) should be used to create a $N \times N$ transformation matrix. After replacing the values for the GHM matrix filter coefficients supplied by the following matrix,

$$\begin{bmatrix} H_{0_{0,0}} & H_{0_{0,1}} & H_{1_{0,0}} & H_{1_{0,1}} & \cdots & \cdots & \cdots & \cdots & \cdots \\ H_{0_{1,0}} & H_{0_{1,1}} & H_{1_{1,0}} & H_{1_{1,1}} & \cdots & \cdots & \cdots & \cdots & \cdots \\ G_{0_{0,0}} & G_{0_{0,1}} & G_{1_{0,0}} & G_{1_{1,0}} & \cdots & \cdots & \cdots & \cdots & \cdots \\ G_{0_{1,0}} & G_{0_{1,1}} & G_{1_{1,0}} & G_{1_{1,1}} & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & H_{0_{0,0}} & H_{0_{0,1}} & H_{1_{0,0}} & H_{1_{0,1}} & \cdots \\ 0 & 0 & 0 & 0 & H_{0_{1,0}} & H_{0_{1,1}} & H_{1_{1,0}} & H_{1_{1,1}} & \cdots \\ 0 & 0 & 0 & 0 & G_{0_{0,0}} & G_{0_{0,1}} & G_{1_{0,0}} & G_{1_{0,1}} & \cdots \\ 0 & 0 & 0 & 0 & G_{0_{1,0}} & G_{0_{1,1}} & G_{1_{1,0}} & G_{1_{1,1}} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

… (11). Preprocessing a $2N \times 2N$ transformation matrix yields the same dimensions as the dimensions of the input matrix after preprocessing.

### 5.4.3. Row Preprocessing

Row preprocessing doubles the number of the input matrix rows. Therefore, if the 2-D input consists of $N \times N$ matrix elements, the outcome of row preprocessing is a $2N \times N$ matrix. The odd rows 1,3,...,2N-1 of this resultant matrix are the same original matrix rows values 1,2,3,..., N, respectively. While the even rows 2,4,...2N represent the initial signal rows multiplied by $\alpha$. For GHM system function $\alpha = 1/\sqrt{2}$.

### 5.4.4. Transforming Rows Can Be Accomplished As Follows

**a.** Apply matrix multiplication to the $2N \times N$ input matrix that has been preprocessed.

**b.** Permute the resulting $2N \times N$ matrix rows by putting the row pairs 1,2 and 5,6..., 2N-3,2N-2, after others in the upper half, then the row pairs 3,4 and 7,8,..., 2N-1,2N below them in the next lower half.

### 5.4.5. Preprocess Columns

to repeat the approach used for preprocessing rows.

**a.** Transpose the $2N \times N$ matrix that was row converted in step 4.

**b.** Repeat step 3 on the $N \times 2N$ matrix (transposition of the row-transformed $2N \times 2N$ matrix), producing a $2N \times 2N$ column preprocessed matrix.

### 5.4.6. The Transformation Column

transformation columns are done to the preprocessed $2N \times 2N$column matrix as follows:

**a**. Perform matrix multiplication on the $2N \times 2N$ constructed transformation matrix produced by the $2N \times 2N$ column preprocessed matrix.

**b.** Change the order of the rows in the resulting $2N \times 2N$ matrix by putting the row pairs 1,2, and 5,6,..., and 2N-3,2N-2 next to each other in the upper half of the rows, then placing the row pairs 3,4, and 7, 8,..., and 2N-1,2N below them in the next lower half.

### 5.4.7 The Completely Transformed Matrix

Use the following to obtain the final transformed matrix:

**a.** Transpose of the matrix produced by the column transformation.

**b.** Use coefficients permutation on the transpose matrix that you just produced. Each of the four basic sub-bands of the resulting transpose matrix is given a coefficients permutation. This means that each sub-band permutes rows and then permutes columns. With repeated row preprocessing, the original $N \times N$matrix is turned into a $2N \times 2N$ DMWT matrix.

### 5.5. Numerical Exampl

In this section, we will use a numerical example for computing 2-D FDMWT using the Separable Method [8].

1. Let X be input 2-D signal: $X = \begin{bmatrix} 16 & 2 & 3 & 13 \\ 5 & 11 & 10 & 8 \\ 9 & 7 & 6 & 12 \\ 4 & 14 & 15 & 1 \end{bmatrix}_{4\times4}$
2. For this 4×4 matrix input 2-D signal, construct an 8×8 transformation matrix W, using Multiwavelet low and high pass filters.
3. Apply row preprocessing to the input matrix X, using repeated row preprocessing, which results in a PR



Waleed A. M. Al-Jawher, Sh. A. Shaban. (2022). Clustering Of Medical Images Using Multiwavelet Transform And K-Means Algorithms. Journal Port Science Research, 5(1), 35–42. https://doi.org/10.36371/port.2022.1.5

ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS
j. port. sci. res.
Journal port Science Research
Available online www.jport.co
Volume 5, No:1. 2022

matrix.

$$X = \begin{bmatrix} 16 & 2 & 3 & 13 \\ 5 & 11 & 10 & 8 \\ 9 & 7 & 6 & 12 \\ 4 & 14 & 15 & 1 \end{bmatrix} \longrightarrow PR = \begin{bmatrix} 16 & 2 & 3 & 13 \\ 11.3137 & 1.414 & 2.1213 & 9.1924 \\ 5 & 11 & 10 & 8 \\ 3.5355 & 7.778 & 7.071 & 5.6569 \\ 9 & 7 & 6 & 12 \\ 6.364 & 4.9497 & 4.2426 & 8.4853 \\ 4 & 14 & 15 & 1 \\ 2.8284 & 9.8995 & 10.6066 & 0.7071 \end{bmatrix}$$

4. Apply row transformation: [Z]=[W] × [PR]

$$Z = W.PR = \begin{bmatrix} 17.9605 & 6.6468 & 7.2125 & 16.2635 \\ 4.0500 & 11.4500 & 9.9500 & 8.5500 \\ 10.6066 & 12.8693 & 12.3037 & 12.3037 \\ 6.5500 & 11.9500 & 13.4500 & 2.0500 \\ -0.9500 & 0.4500 & -0.0500 & 0.5500 \\ 4.879 & -4.4548 & -4.0305 & 3.6062 \\ 2.5500 & -2.0500 & -1.5500 & 1.0500 \\ 6.4347 & -6.8589 & -7.2832 & 7.7075 \end{bmatrix}_{8\times4}$$

5. Apply columns preprocessing
a. Transpose [Z] matrix

$$[Z]^T = \begin{bmatrix} 17.9605 & 4.0500 & 10.6066 & 6.5500 & -0.9500 & 4.8790 & 2.5500 & 6.4347 \\ 6.6468 & 11.4500 & 12.8693 & 11.9500 & 0.4500 & -4.4548 & -2.0500 & -6.8589 \\ 7.2125 & 9.9500 & 12.3037 & 13.4500 & -0.0500 & -4.0305 & -1.5500 & -7.2832 \\ 16.2635 & 8.5500 & 12.3037 & 2.0500 & 0.5500 & 3.6062 & 1.0500 & 7.7075 \end{bmatrix}_{4\times8}$$

b. preprocessing $[Z]^T$ Results in 8x8 column preprocessed matrix [PC].

$$PC = \begin{bmatrix} 17.9605 & 4.0500 & 10.6066 & 6.5500 & -0.9500 & 4.8790 & 2.5500 & 6.4347 \\ 12.7000 & 2.8638 & 7.5000 & 4.6315 & -0.6718 & 3.4500 & 1.8031 & 4.5500 \\ 6.6468 & 11.4500 & 12.8693 & 11.9500 & 0.4500 & -4.4548 & -2.0500 & -6.8589 \\ 4.7000 & 8.0964 & 9.1000 & 8.4499 & 0.3182 & -3.1500 & -1.4496 & -4.8500 \\ 7.2125 & 9.9500 & 12.3037 & 13.4500 & -0.0500 & -4.0305 & -1.5500 & -7.2832 \\ 5.1000 & 7.0357 & 8.7000 & 9.5106 & -0.0354 & -2.8500 & -1.0960 & -5.1500 \\ 16.2635 & 8.5500 & 12.3037 & 2.0500 & 0.5500 & 3.6062 & 1.0500 & 7.7075 \\ 11.5000 & 6.0458 & 8.7000 & 1.4496 & 0.3889 & 2.5500 & 0.7425 & 5.4500 \end{bmatrix}$$

6. 6.Transformation of input columns: Let [BB]=[W] × [PC]

$$BB = \begin{bmatrix} 20.6000 & 8.8671 & 15.9600 & 11.5541 & -0.7495 & 2.9400 & 1.6546 & 3.4600 \\ 4.0729 & 12.6250 & 13.1805 & 13.9750 & 0.5750 & -6.5973 & -2.9750 & -10.3733 \\ 14.0400 & 13.4775 & 17.4000 & 14.1846 & 0.1838 & -2.4600 & -1.0889 & -3.9400 \\ 19.0636 & 6.7750 & 11.7663 & 0.6250 & 0.2250 & 5.9185 & 2.1750 & 11.0521 \\ -2.5739 & 1.1750 & 0.3111 & 2.0250 & 0.1250 & -2.1425 & -0.9250 & -3.5143 \\ 2.7600 & -2.5244 & -0.8400 & -0.1909 & -0.6152 & 2.2500 & 1.2940 & 2.5500 \\ 2.8001 & -1.7750 & -0.5374 & -1.4250 & -0.3250 & 2.3122 & 1.1250 & 3.3446 \\ -1.1600 & -1.7183 & -0.7600 & 4.4336 & -0.7990 & -1.0500 & 0.1202 & -3.7500 \end{bmatrix}$$

7. To get the final transformed matrix

a. Transpose [BB] to get [y].
b. Apply coefficient permutation to the [Y] matrix to get the [YY] matrix.

$$Y = \left[\begin{array}{cccc|cccc} 20.6000 & 4.0729 & 14.0400 & 19.0636 & -2.5739 & 2.7600 & 2.8001 & -1.1600 \\ 8.8671 & 12.6250 & 13.4775 & 6.7750 & 1.1750 & -2.5244 & -1.7750 & -1.7183 \\ 15.9600 & 13.1805 & 17.4000 & 11.7663 & 0.3111 & -0.8400 & -0.5374 & -0.7600 \\ 11.5541 & 13.9750 & 14.1846 & 0.6250 & 2.0250 & -0.1909 & -1.4250 & 4.4336 \\ \hline -0.7495 & 0.5750 & 0.1838 & 0.2250 & 0.1250 & -0.6152 & -0.3250 & -0.7990 \\ 2.9400 & -6.5973 & -2.4600 & 5.9185 & -2.1425 & 2.2500 & 2.3122 & -1.0500 \\ 1.6546 & -2.9750 & -1.0889 & 2.1750 & -0.9250 & 1.2940 & 1.1250 & 0.1202 \\ 3.4600 & -10.3733 & -3.9400 & 11.0521 & -3.5143 & 2.5500 & 3.3446 & -3.75000 \end{array}\right]$$


Waleed A. M. Al-Jawher, Sh. A. Shaban. (2022). Clustering Of Medical Images Using Multiwavelet Transform And K-Means Algorithms. Journal Port Science Research, 5(1), 35–42. https://doi.org/10.36371/port.2022.1.5

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 5, No:1. 2022

$$YY=\left[\begin{array}{cccc:cccc}
20.6000 & 14.0400 & 4.0729 & 19.0636 & -2.5739 & 2.8001 & 2.7600 & -1.1600 \\
15.9600 & 17.4000 & 13.1805 & 11.7663 & 0.3111 & -0.5374 & -0.8400 & -0.7600 \\
8.8671 & 13.4775 & 12.6250 & 6.7750 & 1.1750 & -1.7750 & -2.5244 & -1.7183 \\
11.5541 & 14.1846 & 13.9750 & 0.6250 & 2.0250 & -1.4250 & -0.1909 & 4.4336 \\
\hdashline
-0.7495 & 0.1838 & 0.5750 & 0.2250 & 0.1250 & -0.3250 & -0.6152 & -0.7990 \\
1.6546 & -1.0889 & -2.9750 & 2.1750 & -0.9250 & 1.1250 & 1.2940 & 0.1202 \\
2.9400 & -2.4600 & -6.5973 & 5.9185 & -2.1425 & 2.3122 & 2.2500 & -1.0500 \\
3.4600 & -3.9400 & -10.3733 & 11.0521 & -3.5143 & 3.3446 & 2.5500 & -3.7500
\end{array}\right]$$

## 6. The Proposed Clustering Algorithm

Clustering is the process of automatically organizing data into meaningful groups, or "clusters." Data points, observations, and feature vectors are all terms used to describe individual data bits. The data that is contained within each cluster are equivalent to one another but distinct from the data that are contained within the other clusters. Clustering is a foundational process in many software packages; hence, it must be reliable. Among the many applications of cluster analysis is data compression, the study of climate, physiology, medicine, and business. Essentially, simple image clustering is the method through which collections of images from a database are assembled. Each image in the archive is then assigned a "class label" based on the results of the clustering procedure. Images with the same "class label" are considered to be of the same type. The primary objective of image clustering is to combine existing image databases. [3].

### 6.1 K-Means Clustering

K-MEANS Clustering is a type of cluster analysis that aims to divide n objects into k groups so that each data is placed in the cluster with the closest mean. K-means is the algorithm's name, where k is the number of groups. A case corresponds to the location in the cluster where it is closest to the cluster mean. The algorithm's focus is on achieving k-means. We begin with the first set of criteria and break instances into groups depending on their proximity to the centres. Next, the cluster mean is recalculated using the subjects that are already included in the cluster. Every case is reclassified based on the new set of rules. This phase is done until there is little variation in the cluster meanings from step to step. Finally, we recalculate the cluster averages and place the cases in permanent clusters.

### 6.2 Algorithmic Steps For K-Means Clustering

Let X = {X1, X2, X3…….,Xn} be the set of data points and V = {V1,V2,…….,Vc} be the set of centers.

1. Randomly select 'c' cluster centres.

2. Calculate the distances between each data point and cluster centre.

3. Assign the data point to the cluster centre whose distance from the cluster centre is the minimum of all the cluster centres.

4. Recalculate the new cluster centre using $V_{i=}(1/C_i)\sum_{j=1}^{C_i} X_i$. Where 'ci' represents the number of data points in $i^{th}$ cluster.

5) Recalculate the distance between each data point and the new obtained cluster centres. Distances: Quantitative Variables in K-Means. Identity(absolute) error $d_j\left(x_{ij,}x_{\acute{\imath}j}\right) = I\left(x_{ij\neq}x_{\acute{\imath}j}\right)$. Squared distance $d_j\left(x_{ij,}x_{\acute{\imath}j}\right) = \left(x_{ij}-x_{\acute{\imath}j}\right)^2$

$L_q$ norms $L_{qi\acute{\imath}} = \left[\sum_j \left|x_{ij-}x_{\acute{\imath}j}\right|^q\right]^{1/q}$. Canberra distance $d_{i\acute{\imath}} = \sum_j \frac{|x_{ij}-x_{\acute{\imath}j}|}{|x_{ij}+x_{\acute{\imath}j}|}$. Correlation $p\left(x_{i,}x_{\acute{\imath}}\right) = \frac{\sum_j (x_{ij}-\bar{x}_i)(x_{\acute{\imath}j}-\bar{x}_{\acute{\imath}})}{\sqrt{\sum_j (x_{ij}-\bar{x}_i)^2 \sum_j (x_{\acute{\imath}j}-\bar{x}_{\acute{\imath}})^2}}$

6) Stop if no data point was reassigned; else, repeat from step 3. To divide the dataset into k clusters based on the variance, you must find the k points mj (j = 1, 2,..., k) in Rd that make the distance between xi and mj the smallest. Cluster centroids are the points mj, where j = 1, 2..., k.

• Ordinal variables can be allowed to fall between 0 and 1 so that quantitative metrics can be used:

• Distances between each pair of categories in a categorical variable must be set by the user [2].



Waleed A. M. Al-Jawher, Sh. A. Shaban. (2022). Clustering Of Medical Images Using Multiwavelet Transform And K-Means Algorithms. Journal Port Science Research, 5(1), 35–42. https://doi.org/10.36371/port.2022.1.5

j. port. sci. res.
ISSN: 2616-7441 (Online)
ISSN: 2616-6232 (Print)
ISSN: 2616-7220 (USB)
INTERNATIONAL IDENTIFIER FOR SERIALS

Journal port Science Research
Available online www.jport.co
Volume 5, No:1. 2022

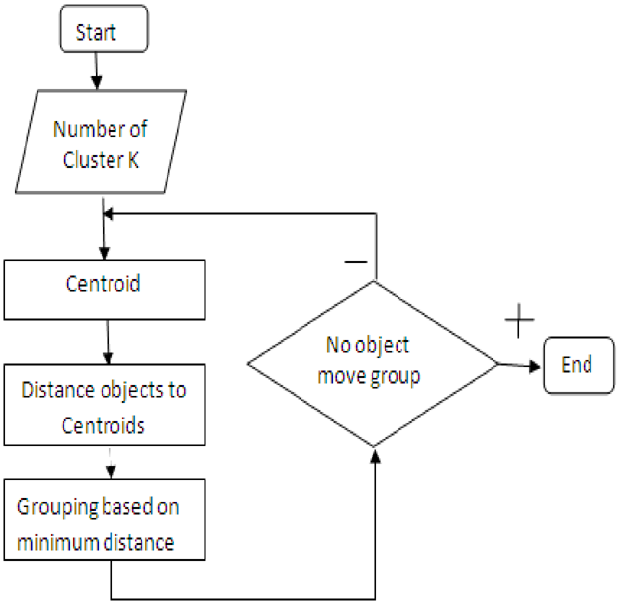


*Fig. (2)* *k-means clustering flowchart*

## 7. PERFORMANCE EVALUATION EXPERIMENTAL RESULTS AND EVALUATION

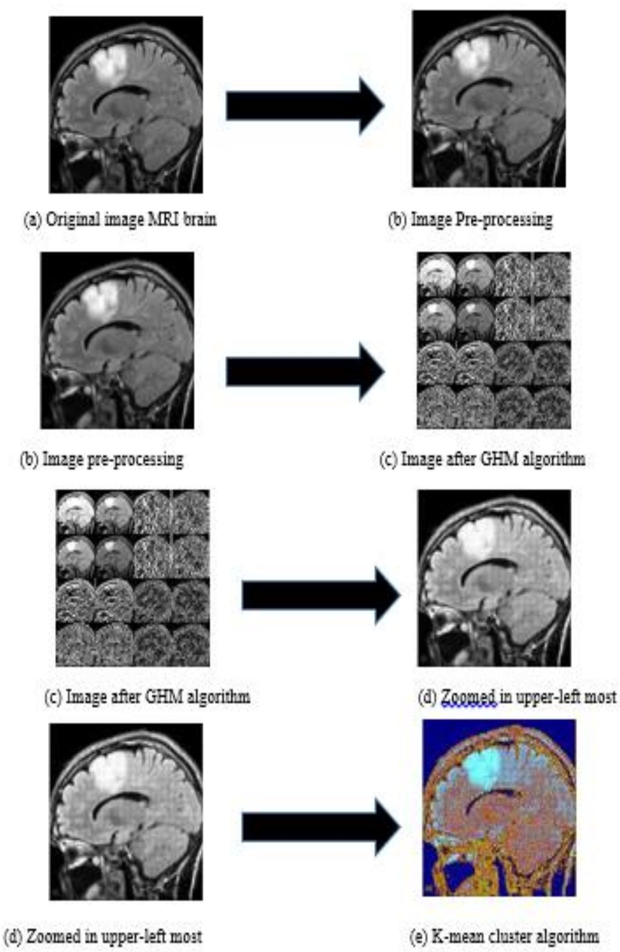


***Fig. (3)*** *The proposed strategy to implement multiwavelet and k-means algorithms. (a) original image without denoising (b) Image after preprocessing (c) image using multiwavelet transform algorithm (d) Image zoomed in upper left most (e) result Image after using GHM algorithm and k-mean algorithm.*

***Table (1).*** *(Evaluation metrics of GHM and k-means algorithm) Table of Images measures calculation using MATLAB software after implementing the Multiwavelet transform algorithm*

| | MSE | PSNR | MNC C | SC | MD | NAE |
|---|---|---|---|---|---|---|
| Image 1 | 2.331 5e+03 | 14.45 45 | 1.367 1 | 0.499 5 | 58 | 0.457 6 |
| Image 2 | 813.4 587 | 19.02 74 | 1.418 5 | 0.461 2 | 60 | 0.511 9 |
| Image 3 | 3.114 9e+03 | 13.19 64 | 1.402 2 | 0.483 9 | 79 | 0.439 2 |
| Image 4 | 474.9 277 | 21.36 45 | 1.414 6 | 0.0.45 97 | 58 | 0.527 4 |
| Image 5 | 7.693 7e+03 | 9.269 5 | 1.932 0 | 0.210 4 | 128 | 1.381 8 |

MSE Mean square Error
PSNR Peak Signal Noise Ratio
MNCC M-Normalized Cross-Correlation
SC Structural Content
MD Maximize Difference
NAE Normalized Absolute Error

## 8. CONCLUSIONS

In this paper, a clustering of medical image data was achieved. The proposed method used the Multiwavelet filterbank transform algorithm and k-means clustering algorithm. It extracts features from MRI medical images and works on them to assist clinicians in diagnosing diseases and detecting the affected part. The proposed method works effectively on the feature extraction due to using the Multiwavelet filterbank transform algorithm. To the best of our knowledge, this is the first study that used the multiwavelet filterbank transform for feature extraction with k-mean clustering. The clustering results and feature extraction can be used as benchmark scores for several new algorithm implementations

## REFERENCES


[1] E. Miranda, M. Aryuni, and E. Irwansyah, "A survey of medical image classification techniques," Proc. 2016 Int. Conf. Inf. Manag. Technol. ICIMTech 2016, no. November, pp. 56–61, 2017, doi: 10.1109/ICIMTech.2016.7930302.

[2] M. Verma, M. Srivastava, N. Chack, A. K. Diswar, and N. Gupta, "A Comparative Study of Various Clustering Algorithms in Data Mining," Int. J. Eng. Res. Appl. www.ijera.com, vol. 2, no. 3, pp. 1379–1384, 2012.

[3] M. Fayez, S. Safwat, and E. Hassanein, "Comparative study of clustering medical images," Proc. 2016 SAI Comput. Conf. SAI 2016, pp. 312–318, 2016, doi: 10.1109/SAI.2016.7556000.

[4] M. Khalid, N. Pal, and K. Arora, "Clustering of Image Data Using K-Means and Fuzzy K-Means," Int. J. Adv. Comput. Sci. Appl., vol. 5, no. 7, pp. 160–163, 2014, doi: 10.14569/ijacsa.2014.050724.

[5] P. Rai and S. Singh, "A Survey of Clustering Techniques," Int. J. Comput. Appl., vol. 7, no. 12, pp. 1–5, 2010, doi: 10.5120/1326-1808.

[6] N. K. Verma, P. Gupta, P. Agrawal, M. Hanmandlu, S. Vasikarla, and Y. Cui, "Medical image segmentation using improved mountain clustering approach," ITNG 2009 - 6th Int. Conf. Inf. Technol. New Gener., pp. 1307–1312, 2009, doi: 10.1109/ITNG.2009.238.

[7] X. W. Li, Y. X. Kang, Y. L. Zhu, G. Zheng, and J. Di Wang, "An improved medical image segmentation algorithm based on clustering techniques," Proc. - 2017 10th Int. Congr. Image Signal Process. Biomed. Eng. Informatics, CISP-BMEI 2017, vol. 2018-Janua, pp. 1–5, 2018, doi: 10.1109/CISP-BMEI.2017.8302178.

[8] Waleed A. Mahmoud Al-Jawher, T Abbas, "Feature combination and mapping using multiwavelet transform" IASJ, AL-Rafidain, Issue 19, Pages 13-34, 2006

[9] A. HM Al-Helali, H. A Ali, B. Al-Dulaimi, D. Alzubaydi & Waleed A Mahmoud "Slantlet transform for multispectral image fusion", Journal of Computer Science, Vol. 5, Issue 4, PP. 263-271, 2009.

[10] AHM Al-Helali, Waleed A. Mahmoud, HA Ali "A Fast personal palm print authentication Based on 3d-multi Wavelet Transformation", TRANSNATIONAL JOURNAL OF SCIENCE AND TECHNOLOGY, Vol. 2, Issue 8, 2012.

[11] H. Al-Taai Waleed A. Mahmoud & M. Abdulwahab "New fast method for computing multiwavelet coefficients from 1D up to 3D", Proc. 1st Int. Conference on Digital Comm. & Comp. App., Jordan, PP. 412-422, 2007.

[12] A H Kattoush, Waleed Ameen Mahmoud Al-Jawher, O Q Al-Thahab "A radon-multiwavelet based OFDM system design and simulation under different channel conditions" Journal of Wireless personal communications, Volume 71, Issue 2, Pages 857-871, 2013.

[13] WA Mahmoud, MS Abdulwahab, HN Al-Taai: "The Determination of 3D Multiwavelet Transform" IJCCCE, vol. 2, issue 4, 2005.

[14] S. A. Lashari and R. Ibrahim, "A Framework for Medical Images Classification Using Soft Set," Procedia Technol., vol. 11, pp. 548–556, 2013, doi: 10.1016/j.protcy.2013.12.227.

[15] D. N. H. Thanh, V. B. S. Prasath, L. M. Hieu, and N. N. Hien, "Melanoma Skin Cancer Detection Method Based on Adaptive Principal Curvature, Colour Normalisation and Feature Extraction with the ABCD Rule," J. Digit. Imaging, vol. 33, no. 3, pp. 574–585, 2020, doi: 10.1007/s10278-019-00316-x.